

Ein Web-Experiment zu Farbharmonie in der Benutzeroberflächengestaltung

Bettina Laugwitz¹

Die Frage nach gültigen ästhetischen Gesetzmäßigkeiten ist prinzipiell für jeden relevant, der sich mit Design befaßt, also auch für Designer von Software-Benutzungs-schnittstellen. In diesem Bereich ist die Gestaltung weitgehend dadurch determiniert, daß die Benutzbarkeit sichergestellt bzw. gar optimiert wird. Benutzbarkeit und Funktionalität sind sicherlich die wesentlichsten Merkmale von Anwendersoftware. Aus der Sicht des Marketings ist es aber auch wichtig, sich (in möglichst positiver Weise) von der Konkurrenz zu unterscheiden, nicht zuletzt durch eine besondere und firmenspezifische Gestaltung der Benutzungsoberfläche (s. z. B. Burmester, Platz, Rudolph & Wild, 1999). Ist ein besonders ästhetisches Design gefragt, können Erkenntnisse aus dem Bereich der empirischen oder auch experimentellen Ästhetik nützlich sein. Man kann sich allerdings zu Recht die Frage stellen, ob die Ergebnisse der Forschung, die zumeist im Labor und mit abstraktem Reizmaterial stattfindet, tatsächlich auf den komplexeren und weniger kontrollierten Anwendungskontext übertragen werden können.

Die referierte Untersuchung befaßt sich in diesem Zusammenhang mit ästhetischer Farbgestaltung. Neben dem Ziel, Eigenschaften ästhetischer Farbkombinationen zu identifizieren bzw. zu bestätigen, soll ermittelt werden, ob die ästhetische Bewertung von Farbkombinationen von einer abstrakten vs. komplex-anwendungsorientierten Darbietungsform abhängt. Darüber hinaus soll der Vergleich der Daten aus dem Labor und dem WWW zeigen, ob die (für den Anwendungskontext durchaus realitätsnahe) mangelnde Kontrolle der Umgebungsbedingungen bei Teilnehmern des Web-Experiments einen systematischen Einfluß auf die Ergebnisse ausübt.

¹ Universität Mannheim, Allgemeine Psychologie (Irtel), 68131 Mannheim
laugwitz@psychologie.uni-mannheim.de. Die Untersuchung fand im Rahmen eines von der Klaus-Tschira-Stiftung gemeinnützige GmbH, Heidelberg, geförderten Projekts statt.

Fragestellungen

Aus der “theory of cognitive hedonics” von Martindale (1984) lassen sich bezüglich der ästhetischen Wirkung von Farbkombinationen verschiedene Hypothesen ableiten. Eine davon sollte in der Untersuchung überprüft werden.

Die Hauptaussage des Modells, das auf einem Netzwerkmodell der Kognition beruht, besteht in der Annahme, daß ein Reiz besonders ästhetisch ist, wenn durch den Prozeß seiner Wahrnehmung eine möglichst große Erregung in den beteiligten Knoten erzeugt wird. Dieser Fall kann nach Martindale z. B. dadurch eintreten, daß die Elemente (hier: Farben) eines Reizes (Farbkombination) sich soweit ähneln, daß die zuständigen Knoten einen höheren Knoten gemeinsam aktivieren, andererseits aber so unterschiedlich sind, daß sie sich nicht gegenseitig hemmen. Daher ist zu vermuten, daß sehr ähnliche Farben in einer Kombination eher negativ bewertet werden und mit zunehmender Unterschiedlichkeit der Komponenten auch die ästhetische Bewertung zunimmt. Nimmt man außerdem an, daß einfarbige Flächen aufgrund der nicht auftretenden Hemmung eine größere Gesamterregung erzeugen als eine Kombination aus sehr ähnlichen Farben, so ist mit einem u-förmigen Zusammenhang zwischen Ähnlichkeit der Farben in der Kombination und *ästhetischer Bewertung der Farbkombination* zu rechnen. Dies wird im Experiment anhand von Kombinationen von Farben untersucht, die sich entweder nur in Hinblick auf Farbton, Helligkeit oder Sättigung unterscheiden. Darüber hinaus wird angenommen, daß die *formale Anordnung* der Farben – hier: Rechtecke vs. Bildschirmmaske – für die Bewertung der Farbkombinationen keine Rolle spielt. Außerdem wird nicht erwartet, daß sich die Beurteilungen der Farbkombinationen durch Versuchspersonen unterscheiden, die am Experiment *im Labor oder über das WWW* teilnehmen.

Methoden

Versuchspersonen: Die 40 Versuchspersonen in den beiden Experimenten im Labor waren z.T. Studierende der Psychologie, z. T. Teilnehmer an den Internationalen Ferienkurse an der Universität Mannheim. An den beiden Web-Experimenten nahmen im Zeitraum vom 18.6. bis zum 24.9.99 72 („Rechtecks-Experiment“) bzw. 64 („Bildschirm-Experiment“) Personen teil. Die Durchführung des Experiments war nur für diejenigen Interessenten möglich, deren System auf die Darstellung einer Farbtiefe von min. 16 Bit eingestellt war. (Die Experimente waren zugänglich über <http://webrum.uni-mannheim.de/ppe/laugwitz/webexp/welcome.htm> sowie über das WWW Experimental Psychology Lab, <http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLabD.html>).

Versuchsanordnung: Die Versuchsdurchführung (Darbietung von Instruktionen und Reizen, Speicherung der Reaktionen der Vpn, der Reaktionszeiten, der demografischen Angaben etc.) wurde gesteuert von einem JAVA-Programm. Im Labor stand dafür ein Rechner mit Monitor, Maus und Tastatur zur Verfügung, auf dem eine lokale Version des Programms zugänglich war. Die Laborexperimente fanden in einem

Raum des Psychophysiklabors der Professur für Allgemeine Psychologie an der Universität Mannheim statt.

Als Reizmaterial dienten Kombinationen aus je drei verschiedenen Farben, die sich entweder in Farbton, Helligkeit oder Sättigung, gemessen in Einheiten des Munsell-Farbsystems, unterschieden. Es gab sieben farbige Bedingungen mit unterschiedlich großem Farbton-Unterschied (incl. der Bedingung ‚Kein Unterschied‘), sechs Bedingungen mit grauen und fünf mit grünen Helligkeitsabstufungen, sowie eine farbige Bedingung mit Sättigungsunterschied (Sättigungs-Unterschied 0 ist identisch mit Farbton-Unterschied 0). Die farbigen Bedingungen waren zum einen dem grünlichen, zum anderen dem rötlichen Bereich des Farbraums entnommen, um zu vermeiden, daß nur die Bewertung der speziellen Farben, nicht aber die der Farbrationen gemessen wurde. Jede Farbkombination wurde in jedem Experiment dreimal dargeboten, wobei die Zuordnung der Farben zu den Teilen des Musters variierte. Daraus ergab sich eine Gesamtanzahl von 81 Trials pro Experiment und Teilnehmer. In einem Telexperiment wurden die Farben innerhalb einer Anordnung von Rechtecken dargeboten, im anderen Telexperiment in Form eines Screenshots einer Bildschirmmaske eines Anwendungsprogramms, deren Elemente entsprechend eingefärbt waren.

Versuchsablauf: Nach der Instruktion erfolgte die Darbietung der Reizmuster auf dem Bildschirm. Unterhalb des Musters waren sieben Radio-Buttons waagrecht angeordnet und mit den Zahlen von -3 bis 3 beschriftet, wobei den äußersten Buttons die Beschriftung ‚gefällt mir überhaupt nicht‘ bzw. ‚gefällt mir sehr gut‘ zugeordnet war. Die Versuchsperson gab ihr Urteil über die Farbkombination ab, indem sie den entsprechenden Button anklickte. In dem Moment verschwand das Muster, und nach einer Pause von 2 Sekunden erschien das nächste. Die Muster wurden in randomisierter Reihenfolge dargeboten. Nach dem letzten Trial erschien ein Formular, in das die Teilnehmer Angaben zu ihrer Person eintragen konnten.

Ergebnisse

Die Daten aus Labor und WWW wurden den Hypothesen entsprechend verschiedenen Varianzanalysen für Meßwiederholungsdesigns unterzogen. An dieser Stelle werden Ergebnisse der Varianzanalysen dargestellt sowie einige inhaltlich interessante Befunde der Nachfolgeanalysen.

Zu den Bedingungen mit unterschiedlichem Farbton: Die Varianzanalyse ergab einen Haupteffekt zulasten des Faktors “Unterschiedlichkeit der Farben” ($F(1272,6) = 59,106, p < .001$), der “Farbklasse” ($F(212,1) = 10,188, p = .002$) sowie eine Wechselwirkung dieser Faktoren ($F(1272,6) = 9,774, p < .001$); außerdem Wechselwirkungen der Unterschiedlichkeit mit den Faktoren “Form der Darbietung” ($F(1272,6) = 6,569, p < .001$) und mit der “Datenquelle” (Web oder Labor; $F(1272,6) = 3,939, p < .001$).

Nachfolgende t-Tests zeigen kurz gefaßt, daß die Bewertung der Unterschiedsstufen 1-3 in positiver Richtung jeweils signifikant von der Kontrollbedingung "Kein Unterschied" abweicht ($t(215) = -5,32, p < .001$; $t(215) = -4,66, p < .001$, $t(215) = -2,24, p = .026$), die der Stufen, 4-6 hingegen in negativer Richtung ($t(215) = 2,38, p = .018$, $t(215) = 4,82, p < .001$, $t(215) = 5,20, p < .001$). Die Wechselwirkungen machen sich dadurch bemerkbar, daß sich die Gruppen 0, 1-3, 4-6 Unterschiedsstufen mehr oder weniger deutlich voneinander trennen; das Grundmuster bleibt aber erhalten.

Zu den Bedingungen mit unterschiedlicher Helligkeit: Sowohl bei den unbunten als auch bei den grünen Kombinationen ist ein signifikanter Haupteffekt zulasten der Unterschiedlichkeit der Farben zu finden ($F(1060,5) = 11,855, p < .001$ bzw. $F(848,4) = 6,075, p < .001$), bei den grünen außerdem eine Wechselwirkung der Unterschiedlichkeit mit der Form der Darbietung ($F(848,4) = 4,855, p = .001$) sowie ein Effekt durch die Quelle der Daten ($F(212,1) = 17,179, p < .001$).

Hier ist ein Anstieg von der einfarbigen Bedingung zu den beiden kleinsten Helligkeitsunterschieden; t-Tests weisen signifikante Unterschiede nach ($t(215) = -4,29, p < .001$ bzw. $t(215) = -4,93, p < .001$); die Unterschiedsstufen 2 bis 5 unterscheiden sich nicht. Bei den grünen Kombinationen gibt es signifikant abweichende Bewertungen für Unterschied 0 und 2 ($t(215) = -2,71, p = .007$) bzw. 0 und 4 ($t(215) = -2,65, p = .009$), wobei auch hier die einfarbige Bedingung negativer beurteilt wird.

Zu den Bedingungen mit unterschiedlicher Sättigung: Die Analyse ergibt Haupteffekte zulasten der Faktoren "Unterschiedlichkeit der Farben" ($F(212,1) = 6,540, p = .011$) und "Farbklasse" ($F(212,1) = 18,658, p < .001$) sowie zulasten der Quelle der Daten ($F(212,1) = 16,236, p < .001$; die Muster schneiden im Labor besser ab als im Web). Die Kombinationen aus unterschiedlich gesättigten Farben werden positiver bewertet als die Muster ohne Sättigungsunterschied ($t(215) = -2,42, p = .016$), und die grünen positiver als die roten ($t(215) = -4,94, p < .001$).

Diskussion

Die Hypothese des Hemmungseffekts im Sinne Martindales (1984) bei ähnlichen Farben muß offensichtlich verworfen werden. Für alle drei Farbmerkmale gilt: kleine Unterschiede sind relativ attraktiv. Bei Kombinationen mit unterschiedlichem Farbton werden sie als gefälliger bewertet als einfarbige Muster und als Muster aus relativ unähnlichen Farben. Inwiefern diese Regel allgemeingültig ist, muß in Anbetracht früherer, tendenziell gegenläufiger Befunde (Laugwitz, 1996, 1999) allerdings fraglich bleiben. Ebenso überraschend ist der analoge Effekt bei Unterschieden in der Sättigung. Ursache der inkonsistenten Ergebnisse könnte sein, daß aus technischen Gründen und praktischen Erwägungen für die aktuellen Experimente relativ helle und ungesättigte Farben ausgewählt wurden.

Für Helligkeitsunterschiede zwischen den Farben in einer Kombination gilt im Einklang mit den Ergebnissen der älteren Experimente, daß sie zu einer positiven

Bewertung führen im Vergleich zu einfarbigen Mustern. Die unbunten Kombinationen erreichen ein Maximum bei mittleren bis starken Unterschieden.

Die Übertragbarkeit von Bewertungen einfacher Muster auf die von komplexeren Mustern, die einem Anwendungsbereich entnommen sind, muß differenziert beurteilt werden. Im großen und ganzen ist die Übereinstimmung der Datenmuster durchaus beachtlich. Dennoch ist es plausibel, daß manche Farben oder auch Kombinationen im Anwendungskontext "Benutzungsoberflächendesign" zumindest befremdlich, wenn nicht gar ungeeignet erscheinen (und dies die Bewertung der *ästhetischen* Erscheinung trotz entsprechender Hinweise in der Instruktion beeinflusst). Dies sollte in der Anwendung auf jeden Fall kontrolliert und berücksichtigt werden, auch wenn die im Experiment gefundenen Wechselwirkungen keine großen Effektstärken aufweisen.

Der Vergleich von Daten von Versuchspersonen aus dem Labor und über das WWW ergibt nur an einem Punkt einen stärkeren Effekt, nämlich beim Haupteffekt bei der Bewertung der grünen Bedingungen zur Helligkeitsvariation. Dies kann als Hinweis darauf gewertet werden, daß die ästhetische Wirkung von Farben, die auf dem Computerbildschirm dargeboten werden, durchaus von der Darstellung des Monitors, möglicherweise auch von anderen Umgebungsbedingungen, abhängt. Es ging in der Untersuchung aber nicht um die Bewertung einzelner Farben, sondern vielmehr um Farbkombinationen, die sich durch bestimmte Relationen zwischen den Farben auszeichnen. Ein entsprechender Haupteffekt zu Lasten der Datenquelle zeigt nur, daß mit den Umgebungsbedingungen der Ankerwert des Ratings schwankt; die relativen Bewertungen der verschiedenen Kombinationen bleibt davon weitgehend unberührt. Die (zwar signifikante, aber effektschwache) Wechselwirkung von Datenquelle und Größe des *Farbton*unterschieds ist kritischer zu betrachten. Da aber die grundlegende Form des Zusammenhangs zwischen Farbtonunterschied und ästhetischer Bewertung erhalten bleibt, hat das Medium WWW in diesem Falle nicht versagt. Vielmehr ist davon auszugehen, daß die Datenerhebung im Web recht gut den Einfluß der Bedingungsunterschiede bei verschiedenen Computernutzern widerspiegelt, und dies würde bedeuten, daß das Web für die Untersuchung von Fragestellungen bezüglich der Farbgestaltung von Benutzeroberflächen ein adäquates Forschungsmedium ist.

Literatur

- Burmester, M., Platz, A., Rudolph, U. & Wild, B. (1999). Aesthetic design – just an add on? In H.-J. Bullinger und Ziegler, J. (eds.), *Human-Computer Interaction: Ergonomics and user interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction* (pp. 671-675). Mahwah: Lawrence Erlbaum.
- Laugwitz, B. (1996). *Experimentelle Überprüfung einer kognitionspsychologischen Theorie der ästhetischen Bewertung von Farbkombinationen*. Unveröffentlichte Diplomarbeit, Universität Mannheim.

- Laugwitz, B. (1999). Der Einfluß von unbunten Flächenbegrenzungen auf die ästhetische Bewertung von Farbkombinationen. In E. Schröger, A. Mecklinger & A. Widmann (Hrsg.), *Experimentelle Psychologie: 41. Tagung experimentell arbeitender Psychologen* (S. 362). Lengerich: Pabst.
- Martindale, C. (1984). The pleasures of thought: A theory of cognitive hedonics. *Journal of Mind and Behavior*, 5 (1), 49-80.