

Der Einsatz von Farbe bei der Gestaltung von Benutzungsoberflächen in der Mensch-Computer Interaktion

Matthias Rauterberg*

Institut für Arbeitspsychologie (IfAP)
Eidgenössische Technische Hochschule (ETH)
Nelkenstr. 11, CH-8092 Zürich, Schweiz

Schlüsselbegriffe: benutzungsorientierter Benchmarktest, Farboberfläche, Softwareergonomie
Keywords: user-oriented benchmark test, color display, software ergonomics

Zusammenfassung

Die Effekte von farbigen Benutzungsoberflächen zur realitätsgerechten Objektdarstellung, zur Transparenz räumlicher Vorstellungen und zur Unterstützung handlungsorientierter Dialogstrukturen wurden mit drei benutzungsorientierten Benchmarktests getestet. Sobald Farbe zur unmittelbaren Handlungssteuerung eingesetzt wird, ergeben sich klar nachweisbare Vorteile für den Einsatz von Farbe bei der Gestaltung von Benutzungsoberflächen. Ergebnisse einer empirischen Vergleichsstudie zeigen, daß eine vollständig freie individuelle Farbwahl zu ergonomisch nicht sinnvollen Konfigurationen führen kann. Umfrageergebnisse zeigen, daß die meisten Benutzer die Standardfarbeinstellung individuell abändern. Ein Regelsatz zur Bewertung der ergonomischen Qualität von farbigen Oberflächen wird vorgestellt. Gestaltungsvorschläge für den Farbeinsatz werden diskutiert.

Summary

To test the effects of color presenting animated objects, increasing the transparency of spatial structures, and supporting action oriented dialogue structures, three experimental investigations were carried out. An advantage of color measured by performance variables was found only, if color was used to support an action oriented dialogue structure. The results of an other experimental investigation show, that in most of the cases an unconstrained individual choice of color configurations leads to ergonomical inappropriate solutions. An inquiry among ETH employees show, that nearly all users of color displays changed the default color configuration settings. To evaluate the ergonomical quality of a color configuration setting an evaluation scheme for color configuration settings was developed and consists of 12 evaluation rules. Design recommendations of color usage in interface programming are discussed.

1. Farbe zur Erhöhung der Transparenz

Farbe läßt sich bei der Gestaltung von Benutzungsoberflächen zu verschiedenen Zwecken verwenden: zur Darstellung von "realitätsgerechten" Anwendungsobjekten, zur Aufmerksamkeitssteuerung bei der Dialoggestaltung, zur Animation von "normalen" Ein/Ausgabeschnittstellen, etc. Es ist abzuklären, in welchen spezifischen Anwendungskontexten Farbe mit welchen Effekten eingesetzt werden kann.

* Der vorliegende Beitrag entstand im Rahmen des Forschungsprojekt BOSS - Benutzer-orientierte Softwareentwicklung und Schnittstellengestaltung (Förderkennzeichen 01 HK 706-0), das vom BMFT (AuT-Programm) gefördert wird.

Ausgehend von der Fülle an positiven Befunden zugunsten von Farbe bei der Suche und Identifikation von statisch vorgegebenem Bildmaterial (Christ 1978) kann angenommen werden, daß bei der Handhabung interaktiver Software diese positiven Effekte von Farbe sich sozusagen von Dialogschritt zu Dialogschritt "aufaddieren". Aus diesem Grunde wurden eine Reihe von benutzungsorientierten Benchmarktests (Rauterberg 1991) durchgeführt. Als arbeitswissenschaftliche Grundlage diente uns das Konzept zur "benutzer-orientierten Dialoggestaltung" von Ulich (1991), wobei im Rahmen dieser Arbeit die beiden Kriterien "Transparenz" und "Individualisierbarkeit" von besonderer Bedeutung sind.

In den folgenden drei Experimenten wird der Frage nachgegangen, welchen Effekt der Einsatz von Farbe bei der Gestaltung von Benutzungsoberflächen auf die "Transparenz" der interaktiven Software hat.

1.1. Farbe zur realitätsgerechten Objektdarstellung bei Simulationsprogrammen

In einer experimentellen Vergleichsstudie (Baumgartner, Bauer & Messmer 1991) wurde versucht herauszufinden, welchen Einfluß die Farbgebung der Anwendungsobjekte (Häuser, Strassen, Flüsse, etc.) beim Umgang mit dem Simulationsprogramm SIMCITY™ (Wright 1989) ausübt. SIMCITY ist ein Computersimulationsspiel, bei dem der Benutzer die Rolle eines Bürgermeisters einnimmt und aufgefordert ist, eine sich entwickelnde Stadt unter "Echtzeit"-Bedingungen zu regieren. Dazu stehen dem Benutzer ein begrenztes Budget sowie eine Reihe von Einflußmöglichkeiten zur Verfügung, deren Anwendung mit unterschiedlichen Kosten verbunden ist. Das Programm steht in zwei Versionen zur Verfügung: monochrom und farbig. Wir stellen die Hypothese auf, daß die Benutzung dieser beiden Versionen unterschiedliche Auswirkungen auf den Spielzustand hat.

Die Untersuchung wurde auf einem Olivetti™-PC durchgeführt (Olivetti M380/C, Monitor: Olivetti color Modell CDU 1431/HA21, CGA-Modus). Es wurden zwei Benutzergruppen mit je 8 Testteilnehmern gebildet. Diese 16 Benutzer wurden gebeten, innerhalb von jeweils zwei Spieldurchgängen (pro Spieldurchgang 30 min.) eine Stadt mit möglichst vielen und zufriedenen Einwohnern aufzubauen. Acht BenutzerInnen (Gruppe-MC: 2 Frauen, 6 Männer; Durchschnittsalter 21.6 ± 1.1 Jahre) begannen mit der monochromen Version von SIMCITY und wechselten dann auf die Farbversion. Die anderen 8 BenutzerInnen (Gruppe-CM: 2 Frauen, 6 Männer; Durchschnittsalter 22.8 ± 1.0 Jahre) fingen mit der Farbversion an und wechselten dann auf die monochrom Version. Bei beiden Spieldurchgängen begannen alle Benutzer wieder bei dem gleichen Startzustand.

Zur Kontrolle wurden folgende Variablen erhoben: Alter, Geschlecht, Farb-/Sehschwäche (Gruppe-MC: 0; Gruppe-CM: 0), EDV-Vorerfahrung hinsichtlich verschiedener Aspekte: Computersysteme (Gruppe-MC: 1091 ± 1107 std.; Gruppe-CM: 1546 ± 1628 std.; $p(t\text{-Test}) \leq .525$), grafische Oberflächen (Gruppe-MC: 826 ± 1187 std.; Gruppe-CM: 735 ± 1220 std.; $p(t\text{-Test}) \leq .882$), monochrom Bildschirme (Gruppe-MC: 412 ± 464 std.; Gruppe-CM: 1156 ± 1473 std.; $p(t\text{-Test}) \leq .209$), color Bildschirme (Gruppe-MC: 504 ± 662 std.; Gruppe-CM: 514 ± 1212 std.; $p(t\text{-Test}) \leq .983$), Computerspiele (Gruppe-MC: 123 ± 146 std.; Gruppe-CM: 178 ± 257 std.; $p(t\text{-Test}) \leq .609$), Simulationsprogramme (Gruppe-MC: 20 ± 16 std.;

Gruppe-CM: 28 ± 41 std.; $p(t\text{-Test}) \leq .619$), SIMCITY (Gruppe-MC: 5.6 ± 12 std.; Gruppe-CM: 0 ± 0 std.), Spielautomaten (Gruppe-MC: 31 ± 59 std.; Gruppe-CM: 13 ± 25 std.; $p(t\text{-Test}) \leq .445$). Zur Erfassung der mentalen Belastung wurde die Eigenzustandsskala von Nitsch (modifiziert nach Apenburg 1986) vor dem ersten Spieldurchgang, zwischen dem ersten und zweiten, sowie nach dem zweiten Spieldurchgang eingesetzt.

Als abhängige Meßgrößen wurden erhoben: drei Variablen zur Charakterisierung des Spielzustandes und 11 fünf-stufige monopolare Ratingskalen aus dem Bereich der subjektiven Beurteilung. Der Spielzustand wurde über das Schätzungsfenster ("evaluation window") zur Halbzeit und am Ende der beiden Spieldurchgänge abgelesen und setzt sich aus den folgenden Variablen zusammen: Anzahl Einwohner ("population"), die öffentliche Meinung ("public opinion"), Nettoübersiedlungsrate ("net migration"). Ein speziell für diese Untersuchung entwickelter, subjektiver Beurteilungsbogen bestehend aus 11 Ratingskalen zu unterschiedlichen Eigenschaften der Ein/Ausgabeschnittstelle, der Dialogstruktur, sowie der Anwendungskomponente wurde nach dem ersten und nach dem zweiten Spieldurchgang vorgelegt.

Bis auf die abhängige Variable "Nettoübersiedlungsrate" ergaben sich für die beiden anderen Meßgrößen des Spielzustandes keine statistisch bedeutsamen Unterschiede zwischen den Testbedingungen. Für die Variable Nettoübersiedlungsrate ("net migration"; positive Werte bedeuten Wachstum der Bevölkerung, negative Werte einen Rückgang der Bevölkerung) wurden folgende Werte gemessen (siehe Tabelle 1).

Tabelle 1: Die Skala "Nettoübersiedlungsrate" des Spielzustandes unterteilt nach den beiden Spieldurchgängen und den beiden Meßzeitpunkten innerhalb jedes Spieldurchganges [Mittelwert \pm Standardabweichung]

Mittelwerte für "Nettoübersiedlungsrate"	1. Spdurchgang		2. Spdurchgang	
	15. Minute	30. Minute	15. Minute	30. Minute
Gruppe-MC	540 ± 879	1055 ± 594	1540 ± 1167	98 ± 640
Gruppe-CM	988 ± 465	478 ± 997	1395 ± 844	1002 ± 1231

Die Mittelwerte für die Variable "Nettoübersiedlungsrate" zusammengefaßt für die beiden Oberflächen (mono: 998 vs. color: 776; siehe Tabelle 2) zeigen einen statistisch nicht als bedeutsam nachweisbaren Vorteil für die monochrome Oberfläche.

Tabelle 2: Die Skala "Nettoübersiedlungsrate" des Spielzustandes unterteilt nach den beiden Meßzeitpunkten innerhalb jedes Spieldurchganges [Gruppen-Mittelwert \pm Standardabweichung]

Mittelwerte für ...	15. Minute	30. Minute	insgesamt
monochrome Oberfläche	967 \pm 942	1029 \pm 934	998
color Oberfläche	1234 \pm 904	288 \pm 832	776
insgesamt	1116	658	887

Die 3-faktorielle Varianzanalyse mit Messwiederholung auf dem Faktor "Zeitpunkt" ergibt nur einen signifikanten Unterschied für den Faktor "Zeitpunkt"; dies heißt, daß alle Testteilnehmer nach der 30. Spielminute eine signifikant geringere Nettoübersiedlung für die von ihnen "regierte" Stadt aufweisen als nach der 15. Spielminute.

Tabelle 3: Ergebnisse der drei-faktoriellen Varianzanalyse (MANOVA der SPSS PC-Version 3.0) bezüglich der Mittelwerte der Skala "Nettoübersiedlungsrate" des Spielzustandes mit Meßwiederholung auf dem Faktor "Zeitpunkt"

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	0.94	0.349
Gruppe	(MC vs CM)	1	0.55	0.471
Oberfläche x Gruppe		1	1.13	0.307
Zeitpunkt	(15.min vs 30.min)	1	6.62	0.022
Gruppe x Zeitpunkt		1	0.01	0.972
Oberfläche x Zeitpunkt		1	3.93	0.670
Oberfläche x Gruppe x Zeitpunkt		1	3.09	0.101

Es ergibt sich somit kein statistisch nachweisbarer Vorteil bei den Performanzmaßen für eine der beiden Benutzungsoberflächen (Tabelle 3). Anders sieht es jedoch bei den Schätzwerten für die subjektive Beurteilung aus. Hier zeigen sich bei drei der 11 Ratingskalen eine statistisch bedeutsame Überlegenheit zugunsten der Farboberfläche (Tabelle 4).

Tabelle 4: Ergebnisse des t-Tests bezüglich der Mittelwerte der drei ausgewählten Skalen des subjektiven Beurteilungsbogen (Skalenrange: 1="stimmt nicht" ... 5="stimmt völlig")

Skalen	"monochrom"	"color"	df	F	p
"Das Spielfeld war übersichtlich (die Objekte waren gut zu erkennen)"	2.7	4.4	1	42.9	0.001
"Ich hatte keine Probleme beim Suchen bestimmter Objekte (zB. Kraftwerk, Wohnzonen, etc.)"	2.9	4.3	1	15.4	0.002
"Die Zonen liessen sich klar unterscheiden (Wohnzonen, Geschäfts-, Industriezonen)"	3.5	4.5	1	6.82	0.021

Für die folgenden acht subjektiven Beurteilungsskalen ergaben sich keine bedeutsamen Mittelwertsunterschiede zwischen der monochromen und der color Oberfläche:

- "In den Wohnzonen gab es die unterschiedlichsten Gebäude";
- "Die Stadt hat sich unkontrolliert entwickelt, da ich den Überblick verloren habe";
- "Auf die Probleme wurde ich durch die Warnmeldungen aufmerksam gemacht";
- "Das Kartenfenster (map window) half mir die dringenden Probleme zu erkennen";
- "Ich konnte mich anhand der Übersichtskarten gut auf dem Spielfeld orientieren";
- "Ich wußte nicht, wie ich die Probleme bekämpfen sollte";
- "Das Spiel langweilte mich";
- "Ich komme mit dem Spiel nach diesem Versuch gut zurecht".

Für den Bereich der mentalen Belastung zeigt die Skala "Anstrengungsbereitschaft" einen signifikanten Effekt. Für die anderen Skalen ("Erholtheit", "Kontaktbereitschaft", "Schläfrigkeit", "soziale Anerkennung" und "Spannungslage") zeigen sich keine Unterschiede zwischen den Untersuchungsbedingungen. Die Skala "Anstrengungsbereitschaft" setzt sich additiv aus den sechs Items "aktiv", "anstrengungsbereit", "arbeitsfreudig", "ausdauernd", "energiegeladen" und "konzentrationsfähig" zusammen, wobei jedes Item mit 1 ("trifft kaum zu") bis 6 ("trifft völlig zu") beantwortet werden kann.

Tabelle 5: Die Skala "Anstrengungsbereitschaft" aus dem Eigenzustandsbogen unterteilt nach den drei Messzeitpunkten und den beiden Gruppen (Skalenrange: 6="trifft kaum zu" bis 36="trifft völlig zu")

Mittelwert der Ausgangslage	nach dem 1. Spiel-		nach dem 2. Spiel-	
...	e	durchgang	durchgang	
Gruppe-MC	20.4 ±4.3	23.5 ±4.7	monochrom	20.4 ±7.4 color
Gruppe-CM	23.1 ±6.0	23.2 ±5.6	color	19.6 ±7.9 monochrom

Die Differenzwerte für die beiden Spieldurchgänge zeigen (gemäß Tabelle 5), daß die Anstrengungsbereitschaft der Gruppe-MC nach dem ersten Spieldurchgang zunimmt (+3.1), während dies bei der Gruppe-CM nicht der Fall ist (+0.1). Obwohl man zunächst annehmen könnte, daß durch die Aufgabebearbeitung die Anstrengungsbereitschaft abnehmen müßte, läßt sich der oft beobachtbare, umgekehrte Effekt dadurch erklären, daß die meisten BenutzerInnen vor Beginn des Experimentes in der Regel aufgeregt sind, weil sie nicht einschätzen können, was im einzelnen auf sie zukommt. Nach der ersten Aufgabebearbeitungsphase spiegelt sich dann die Erleichterung der BenutzerInnen in einem Anstieg der Skalenwerte des Eigenzustandsbogens wider. Nach dem zweiten Spieldurchgang nimmt dann erwartungsgemäß bei beiden Gruppen die Anstrengungsbereitschaft ab.

Die Varianzanalyse mit Messwiederholung auf dem Faktor "Gruppe" ergibt eine signifikante Wechselwirkung "Oberfläche x Gruppe" (Tabelle 6), welche auf den praktisch fehlenden Anstieg der Anstrengungsbereitschaft nach dem ersten Spieldurchgang der Gruppe-CM rückführbar ist.

Tabelle 6: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Anstrengungsbereitschaft" des Eigenzustandsbogens ergibt folgendes Bild

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	0.12	0.733
Gruppe	(MC vs CM)	1	2.12	0.169
Oberfläche x Gruppe		1	6.24	0.027

Ein klar nachweisbarer Vorteil der Farbversion tritt also nur bei der subjektiven Beurteilung auf. Dieses Ergebnis steht in Übereinstimmung mit den Beobachtungen von Munns (1968). Die Tatsache, daß sich die subjektiv erlebten Vorteile der Farboberfläche nicht auch im Verhaltensbereich aufweisen lassen, kann darin begründet sein, daß die abhängigen Meßgrößen einer zu starken Störvarianz seitens des komplexen Simulationsprogrammes unterworfen sind.

Um den Schwierigkeiten der Meßgenauigkeit der Performanzparameter bei einem komplexen Simulationsprogramm Rechnung zu tragen, wurde ein weiteres Experiment mit dem gleichen Testdesign und der gleichen Hypothese durchgeführt; diesmal jedoch mit dem konzeptionell einfachen Geschicklichkeitsspiel TETRIS, welches primär auf der schnellen adäquaten Wahrnehmung zweidimensionaler Muster basiert.

1.2 Farbe zur Unterstützung benutzergerechter räumlicher Vorstellungen

In dieser experimentellen Vergleichsstudie (Loeliger et al. 1991) wurden die monochrom Version des Computerspiels TETRIS™ (Pazhitnov 1988) mit der farbigen Oberfläche von TETRIS verglichen. Bei diesem Geschicklichkeitsspiel muß der Spieler in einem Schacht die von der "Decke fallenden"

Spielsteine während ihres "Fluges" zum Boden so drehen und/oder kippen, daß sie sich möglichst in die Lücken zwischen den schon im Schacht befindlichen Spielsteinen einpassen. Dazu stehen dem Benutzer vier verschiedene Tasten zur Verfügung: nach links drehen, nach rechts drehen, um die x-Achse kippen und um die y-Achse kippen. Wenn eine "Zeile" quer über die gesamte Schachtbreite vollständig mit Spielsteinflächen ausgefüllt ist, verschwindet diese "Zeile" automatisch. Ziel des Spieles ist es, möglichst viele "Zeilen" zu erzeugen.

Wir entschieden uns für die englisch-sprachige, zweidimensionale Darstellung der einzelnen Spielsteine, bei der die Farbe redundant zu der Form der einzelnen Spielsteine codiert ist: 1. der geometrische Typ des einzelnen Spielsteins wird eindeutig durch seine Form charakterisiert, 2. durch sein Muster und 3. zusätzlich bei der Farbversion eindeutig über seine Farbe. Wir stellen wiederum die Hypothese auf, daß die Benutzung der beiden Versionen unterschiedliche Auswirkungen auf die "Spielgüte" hat.

Die Untersuchung wurde auf einem MacintoshTM-IIx durchgeführt (Apple Color High-Resolution RGB Monitor Modell MO401Z; 256 Farben). Es wurden wiederum 16 BenutzerInnen gebeten, innerhalb von insgesamt zwei Spieldurchläufen möglichst viele "vollständige Bodenbeläge" (engl. "lines") herzustellen. Jede Spielphase bestand aus drei Spieldurchgängen von jeweils 5 Minuten Spieldauer, welche sich durch aufsteigend schwieriger werdende Spielstufen (engl. "level": 1, 3 und 5) auszeichneten. Die unterschiedliche Schwierigkeitsstufe wird über unterschiedliche "Fallgeschwindigkeiten" erzeugt. Innerhalb der 5 Minuten eines Spieldurchganges konnten die BenutzerInnen beliebig oft mit dem Spiel neu beginnen. Gruppe-MC bestand aus 8 BenutzerInnen (3 Frauen, 5 Männer; 22.8 ± 1.1 Jahre), welche zuerst mit der monochrom Version (3 Spieldurchgänge à 5 min.) begannen und dann auf die Farbversion (3 Spieldurchgänge à 5 min.) wechselten. Gruppe-CM bestand ebenfalls aus 8 BenutzerInnen (5 Frauen, 3 Männer; 22.0 ± 0.8 Jahre), welche mit der Farbversion anfangen und dann zu der monochrom Version übergangen. Bei allen 6 Spieldurchgängen begann jeder Benutzer wieder mit einem "leeren Schacht".

Zur Kontrolle wurden folgende Variablen erhoben: Alter, Geschlecht, EDV-Vorerfahrung hinsichtlich verschiedener Aspekte: Computersysteme (Gruppe-MC: 602 ± 1198 std.; Gruppe-CM: 196 ± 252 std.; $p(t\text{-Test}) \leq .380$), Programmiersprachen (Gruppe-MC: 604 ± 1198 std.; Gruppe-CM: 206 ± 229 std.; $p(t\text{-Test}) \leq .390$), grafische Oberflächen (Gruppe-MC: 307 ± 503 std.; Gruppe-CM: 275 ± 302 std.; $p(t\text{-Test}) \leq .880$), Computerspiele (Gruppe-MC: 39 ± 41 std.; Gruppe-CM: 71 ± 117 std.; $p(t\text{-Test}) \leq .480$), TETRIS (Gruppe-MC: 3.8 ± 5.2 std.; Gruppe-CM: 2.5 ± 4.6 std.; $p(t\text{-Test}) \leq .620$). Zur Erfassung der mentalen Belastung wurde die Eigenzustandsskala von Nitsch (modifiziert nach Apenburg 1986) vor der ersten Spielphase ("Ausgangslage"), zwischen der ersten und zweiten, sowie nach der zweiten Spielphase eingesetzt.

Als abhängige Meßgrößen wurden erhoben: der Spielzustand als Performanzmaß ("number of lines") und die subjektive Beurteilung nach jeder Spielphase. Der für dieses Spiel entwickelte subjektive Beurteilungsbogen bestand aus 7 fünf-stufigen monopolaren Ratingskalen zu unterschiedlichen Eigenschaften der Ein/Ausgabeschnittstelle, der Dialogstruktur, sowie der Anwendungskomponente und wurde nach der ersten und nach der zweiten Spielphase von den BenutzerInnen ausgefüllt.

Bei dem Performanzmaß "Anzahl vollständige Zeilen" ergibt sich kein statistisch bedeutsamer Vorteil für eine der beiden Oberflächen. Es läßt sich jedoch eine signifikante Wechselwirkung "Oberfläche x Gruppe" finden (siehe Tabelle 9). Diese Wechselwirkung bedeutet, daß sich zwar die beiden Gruppen (Gruppe-MC durchschnittlich 11.6 Zeilen und Gruppe-CM durchschnittlich 12.6 Zeilen bei "mono") während der jeweiligen Spielphase mit der monochromen Oberfläche praktisch nicht unterscheiden, sich aber die Gruppe-MC in der Spielphase mit der farbigen Oberfläche (durchschnittlich 16.0 Zeilen bei "color") deutlich von der Gruppe-CM (durchschnittlich 11.6 Zeilen bei "color") abhebt, so alsoob die BenutzerInnen der Gruppe-MC beim Übergang von der monochromen zur farbigen Oberfläche einen Aktivitätsschub erfahren haben. Die damit einhergehende Beanspruchung spiegelt sich dann auch in den Werten des Eigenzustandsbogens wider (siehe Tabelle 11).

Tabelle 7: Die Skala "Anzahl vollständige Zeilen" des Spielzustandes unterteilt nach den beiden Spielphasen und den drei Spielstufen ("Schwierigkeitsgrad": 1=leicht, 3=mittel, 5=schwer) [Mittelwert \pm Standardabweichung]

Mittelwerte für "Anzahl Zeilen"	1. Spielphase			2. Spielphase		
	Stufe-	Stufe-	Stufe-	Stufe-	Stufe-	Stufe-
Gruppe-MC	10.5 \pm 6.2	10.6 \pm 6.8	13.8 \pm 5.9	16.8 \pm 4.7	15.3 \pm 7.1	16.0 \pm 9.6
Gruppe-CM	9.9 \pm 7.0	11.1 \pm 8.5	13.8 \pm 8.1	12.8 \pm 7.2	13.5 \pm 7.0	12.1 \pm 7.5

Tabelle 8: Die Mittelwerte (\pm Standardabweichung) der Variablen "Anzahl vollständige Zeilen" zusammengefaßt nach Oberflächentyp untergliedert nach Spielstufen ("Schwierigkeitsgrad": 1=leicht, 3=mittel, 5=schwer)

Mittelwerte für ...	Stufe-1	Stufe-3	Stufe-5	insgesamt
monochrome Oberfläche	11.6 \pm 6.6	12.1 \pm 6.8	12.9 \pm 6.6	12.2
color Oberfläche	13.3 \pm 6.8	13.2 \pm 7.8	14.9 \pm 8.7	13.8
insgesamt	12.5	12.6	13.9	13.0

Tabelle 9: Ergebnisse der drei-faktoriellen Varianzanalyse (MANOVA) bezüglich der Mittelwerte der Skala "Anzahl vollständige Zeilen" des Spielzustandes mit Meßwiederholung auf dem Faktor "Spielstufe"

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	2.64	0.126
Gruppe	(MC vs CM)	1	0.29	0.598
Oberfläche x Gruppe		1	8.21	0.012
Spielstufe	(1 vs 3 vs 5)	2	1.08	0.354
Gruppe x Spielstufe		2	0.34	0.714
Oberfläche x Spielstufe		2	0.08	0.927
Oberfläche x Gruppe x Spielstufe		2	2.15	0.135

Es ergab sich somit kein statistisch nachweisbarer Vorteil bei dem Performanzparameter für die Farbversion. Der nachweisbare Vorteil der Farbversion tritt jedoch wiederum bei der subjektiven Beurteilung auf (siehe Tabelle 10).

Tabelle 10: Ergebnisse der t-Tests bezüglich der Skalenmittelwerte des subjektiven Beurteilungsbogen [in den Spalten "monochrom" und "color" stehen die zugehörigen Mittelwerte]

(Skalenrange: 1="stimmt nicht" ... 5="stimmt völlig")

Skalen	"monochrom"	"color"	df	F	p
"Das Spielfeld ist übersichtlich"	4.2±0.7	4.5±0.5	1	4.1	0.063
"Ich konnte die Form der Spielsteine schnell erkennen"	4.1±0.7	4.6±0.5	1	5.4	0.035

Für die folgenden fünf subjektiven Beurteilungsskalen ergaben sich keine bedeutsamen Mittelwertsunterschiede zwischen der monochromen und der color Oberfläche:

- "Die Fallgeschwindigkeit der Bausteine war etwas zu hoch";
- "Mit der Bedienung der Tastatur hatte ich Mühe";
- "Das Drehen der Spielsteine fand ich eine nützliche Eigenschaft beim Plazieren";
- "Das Produzieren sogenannter 'lines' war einfach";
- "Die 'space'-Taste (Spielsteine fallen lassen) habe ich viel benutzt".

Die Differenzwerte für die Skala "Anstrengungsbereitschaft" des Eigenzustandsbogens unterteilt nach den beiden Spielphasen und den beiden Gruppen ergeben ein ähnliches Bild wie bei dem vorherigen Experiment (vergleiche Tabelle 11 und 6).

Tabelle 11: Die Skala "Anstrengungsbereitschaft" aus dem Eigenzustandsbogen unterteilt nach den beiden Gruppen [gemittelte Differenzwerte \pm Standardabweichung]

Differenzwerte der .für die 1. Spielphase für die 2. Spielpha				
Gruppe-MC	1.0 \pm 2.1	monochrom	-1.6 \pm 2.0	color
Gruppe-CM	0.4 \pm 2.3	color	-0.6 \pm 2.2	monochrom

Die Anstrengungsbereitschaft Gruppe-MC nimmt wiederum während der ersten Spielphase stärker zu (+1.0) als die der Gruppe-CM (+0.4). Nach der zweiten Spielphase nimmt die Anstrengungsbereitschaft für die Gruppe-MC ("color") jedoch stärker ab (-1.6), als diejenige der Gruppe-CM ("monochrom") (-0.6). Dies bedeutet, daß sich die TestteilnehmerInnen mit dem Farbbildschirm nach der zweiten Spielphase stärker beansprucht fühlten, als diejenigen mit dem monochrom Bildschirm. Diese Interpretation ergibt sich aus der 2-faktoriellen Varianzanalyse, bei der sich wie beim vorherigen Experiment eine signifikante Wechselwirkung "Oberfläche x Gruppe" zeigt (siehe Tabelle 12).

Tabelle 12: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Anstrengungsbereitschaft" aus dem Eigenzustandsbogen

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	1.10	0.310
Gruppe	(MC vs CM)	1	0.06	0.810
Oberfläche x Gruppe		1	5.50	0.030

Ein recht ähnliches Bild ergibt sich auch für die beiden Skalen "Spannungslage" und "Stimmungslage" des Eigenzustandsbogens (siehe Tabelle 13 und Tabelle 14).

Tabelle 13: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Spannungslage" aus dem Eigenzustandsbogen

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	1.05	0.323
Gruppe	(MC vs CM)	1	0.01	0.960
Oberfläche x Gruppe		1	3.49	0.083

Tabelle 14: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Stimmungslage" aus dem Eigenzustandsbogen

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	0.03	0.860
Gruppe	(MC vs CM)	1	0.19	0.670
Oberfläche x Gruppe		1	6.41	0.020

Offenbar ist der messbare Effekt im Performanzbereich - bedingt durch Farbe - so klein, daß er mit einer Stichprobengröße von 16 Benutzern und dem hier gewählten Untersuchungszeitraum nicht nachgewiesen werden kann. Eine andere Interpretation der Ergebnisse kann darin gesehen werden, daß die Verwendung von Farbe bei den beiden bisher vorgestellten Experimenten nicht unmittelbar die interaktive Performanz beeinflusst. Um diesem Aspekt nachgehen zu können, wurde das folgende Experiment durchgeführt. Wir fragen uns daher, ob die Verwendung von Farbe – ausschließlich zur Erhöhung der Transparenz der maus-sensitiven Bereiche – einen Einfluß auf die Performanz hat.

1.3 Farbe zur Unterstützung handlungsorientierter Dialogstrukturen

Ein HyperCard™-Stapel (HyperCard 1.01) über "ornitologisches Wissen" (Lynch 1988; englische Fassung) wurde nach SuperCard™ 1.5 portiert, um ihn mit Farbe ausstatten zu können (Cotting & Kohli 1991). Dieser Stapel hat eine HyperText-Benutzungsoberfläche. Alle maus-sensitiven Buttons lassen sich bei der monochrom Version (HyperCard) mit der Tastenkombination "command+option" sichtbar machen. Bei der Farb-Version (SuperCard) wurden dagegen alle semantisch zusammengehörigen maus-sensitiven Bereiche ("buttons") mit demselben, transparenten Farbton und nicht zusammengehörende Bereiche mit unterschiedlichen Farben versehen. Hierdurch wurde erreicht, daß der Benutzer unmittelbar im aktuellen Dialogkontext alle möglichen, semantisch zusammenhängenden Dialogfortsetzungsmöglichkeiten handlungsleitend wahrnehmen konnte (siehe Abbildung 1).

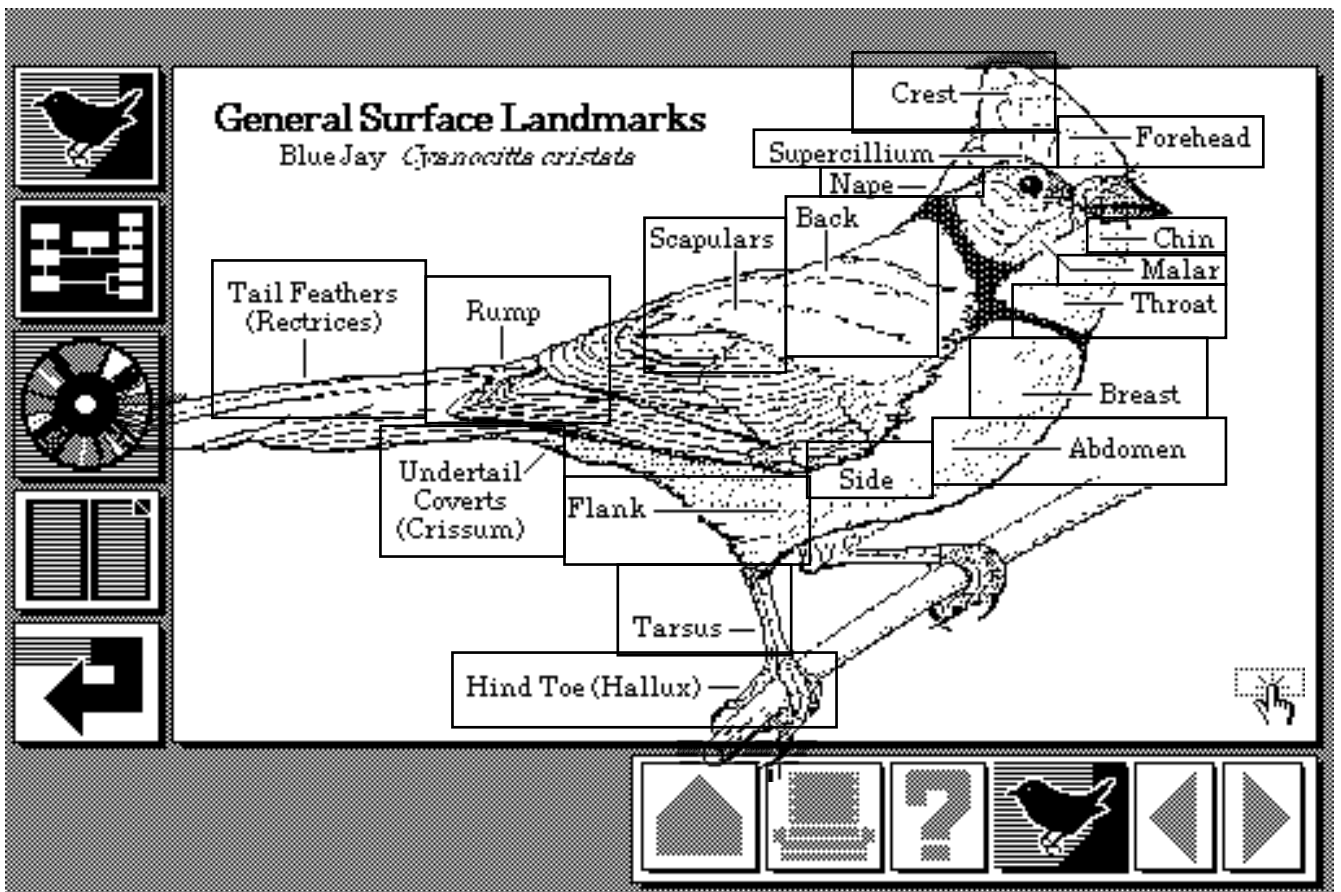


Abbildung 1: Darstellung einer "Karte" des SuperCard-Stacks über ornitologisches Wissen (entnommen aus Lynch 1988); alle maus-sensitiven Bereiche (Rechtecke, "buttons", Ikonen) im dargestellten Motiv sind mit jeweils unterschiedlich gefärbten Rechtecken gekennzeichnet.

In dieser experimentellen Untersuchung wurden 12 BenutzerInnen gebeten, innerhalb von insgesamt zwei Aufgabenserien mit jeweils maximal 20 Fragen möglichst viele dieser Fragen über das ornitologische Wissensgebiet zu beantworten. Für jede Aufgabenserie standen den BenutzerInnen maximal 20 Minuten zur Verfügung. Die 6 Benutzer der Gruppe-MC (0 Frauen, 6 Männer; Durchschnittsalter 24.0 ± 2.4 Jahre) begannen mit der monochrom Version und wechselten dann auf die Farb-Version. Die 6 BenutzerInnen der Gruppe-CM (1 Frau, 5 Männer; Durchschnittsalter 24.0 ± 2.1 Jahre) fingen mit der Farb-Version an und wechselten dann auf die monochrom Version. Um den Lerneffekt bei der wiederholten Bearbeitung zu verringern, wurden zwei parallele, d.h. möglichst ähnliche Aufgabenserien zusammengestellt.

Zur Kontrolle wurden folgende Variablen erhoben: Alter, Geschlecht, EDV-Vorerfahrung hinsichtlich verschiedener Aspekte: Computersysteme (Gruppe-MC: 1908 ± 1921 std.; Gruppe-CM: 839 ± 1329 std.; $p(t\text{-Test}) \leq .289$), Programmiersprachen (Gruppe-MC: 2254 ± 2198 std.; Gruppe-CM: 1333 ± 1200 std.; $p(t\text{-Test}) \leq .389$), grafische Oberflächen (Gruppe-MC: 1862 ± 1968 std.; Gruppe-CM: 1087 ± 1281 std.; $p(t\text{-Test}) \leq .438$), Hyper-/SuperCard (Gruppe-MC: 27 ± 50 std.; Gruppe-CM: 28 ± 49 std.; $p(t\text{-Test}) \leq .955$), Englischkenntnisse (Skala 1="kaum" bis 5="sehr gut": Gruppe-MC: 3.8 ± 0.8 ; Gruppe-CM: 3.3 ± 0.8 ; $p(t\text{-Test}) \leq .296$), ornitologisches Wissen (Skala 1="kaum" bis 5="sehr gut": Gruppe-

MC: 1.7 ± 0.5 ; Gruppe-CM: 1.3 ± 0.5 ; $p(t\text{-Test}) \leq .290$), sowie Farb-/Sehschwäche (Gruppe-MC: 0 Personen; Gruppe-CM: 0 Personen). Zur Erfassung der mentalen Belastung wurde die Eigenzustandsskala nach Nitsch (modifiziert von Apenburg 1986) vor der ersten Aufgabenserie, zwischen der ersten und zweiten, sowie nach der zweiten Aufgabenserie eingesetzt.

Als abhängige Meßgrößen wurden erhoben: die Anzahl der gelösten Aufgaben (= richtig beantwortete Wissensfragen) als Performanzmaß und die subjektive Beurteilung nach jeder Aufgabenserie mittels eines Fragebogens (aus Spinax 1987), welcher aus 11 bipolaren Ratingskalen in Form eines semantischen Differentials bestand (z.B. "übersichtlich": -2, -1, 0, +1, +2:"unübersichtlich", "verwirrend": -2, -1, 0, +1, +2:"eindeutig", etc.; Rauterberg 1991).

Tabelle 15: Die Skala "Anzahl gelöster Aufgaben" unterteilt nach den beiden Benutzer-Gruppen [Mittelwert \pm Standardabweichung]

Mittelwerte für ...	monochrom Oberfläche	color Oberfläche	insgesamt
Gruppe-MC	9.3 ± 2.1	18.2 ± 2.8	13.7 ± 5.2
Gruppe-CM	13.2 ± 4.0	18.0 ± 3.2	15.7 ± 4.2

Bei dem Performanzmaß "Anzahl gelöster Aufgaben" ergibt sich ein Mittelwertsunterschied zugunsten der farbigen Benutzungsoberfläche (siehe Tabelle 15). Dieser Mittelwertsunterschied ist statistisch bedeutsam zugunsten der farbigen Benutzungsoberfläche (siehe Tabelle 16).

Tabelle 16: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Anzahl gelöster Aufgaben"

Faktoren		df	F	p
Oberfläche	(monochrom vs color)	1	8.04	0.018
Gruppe	(MC vs CM)	1	1.27	0.287
Oberfläche x Gruppe		1	93.91	0.001

Zusätzlich zu dem Mittelwertsunterschied der abhängigen Variablen "Anzahl gelöster Aufgaben" ergibt sich eine hoch signifikante Wechselwirkung "Oberfläche x Gruppe" (siehe Tabelle 16), welche darauf zurückzuführen ist, daß die Gruppe-CM bei der ersten Aufgabenserie ("color" Bedingung mit durchschnittlich 18.0 gelösten Aufgaben) der Gruppe-MC ("mono" Bedingung mit durchschnittlich 9.3 gelösten Aufgaben) deutlich überlegen ist (siehe Tabelle 15).

Zusätzlich ergeben sich auch nachweisbare Vorteile zugunsten der Farbversion bei der subjektiven Beurteilung (siehe Tabelle 17).

Tabelle 17: Ergebnisse des t-Tests bezüglich der Mittelwerte der bipolaren Skalen des subjektiven Beurteilungsbogen [in den Spalten "monochrom" und "color" stehen die zugehörigen Mittelwerte] (Skalenrange: -2="trifft zu" ... 0="teils/teils" ... +2="trifft zu")

Skalen des subjektiven Beurteilungsbogen	"monochrom"	"color"	df	F	p
"verwirrend" <-> "eindeutig"	+0.1 ± 0.8	+0.9 ± 0.8	1	14.7	0.003
"unübersichtlich" <-> "übersichtlich"	-0.2 ± 1.2	+0.6 ± 1.2	1	3.7	0.083
"vorgeschrieben" <-> "frei"	-0.4 ± 1.2	+0.3 ± 1.2	1	3.7	0.083

Die Erholtheit beider Gruppen nimmt nach der ersten Aufgabenserie zu. Die Erholtheit für die Gruppe-CM nimmt jedoch nach der zweiten Aufgabenserie ("mono") stärker ab, als diejenige der Gruppe-MC unter der Testbedingung "color" (siehe Tabelle 18). Dies bedeutet, daß sich die BenutzerInnen nach dem Wechsel vom Farbbildschirm auf den monochromen Bildschirm stärker beansprucht fühlten als umgekehrt. Diese Interpretation ergibt sich aus der 2-faktoriellen Varianzanalyse, bei der sich eine signifikante Wechselwirkung "Oberfläche x Gruppe" zeigt (siehe Tabelle 19).

Tabelle 18: Die Differenzwerte für die Skala "Erholtheit" aus dem Eigenzustandsbogen unterteilt nach den beiden Aufgabenserien und den beiden Gruppen [Skalenmittelwert ± Standardabweichung]

Differenzwerte der	für die 1. Aufgabenserie	für die 2. Aufgabenserie
Gruppe-MC	+1.2 ± 3.7 monochrom	-1.2 ± 3.5 color
Gruppe-CM	+2.5 ± 2.5 color	-2.8 ± 2.3 monochrom

Tabelle 19: Der varianzanalytische Test (MANOVA) der Mittelwerte der Skala "Erholtheit" aus dem Eigenzustandsbogen

Faktor	df	F	p
Oberfläche (monochrom vs color)	1	0.03	0.866
Gruppe (MC vs CM)	1	0.06	0.810
Oberfläche x Gruppe	1	6.66	0.027

Bei allen anderen Skalen des Eigenzustandsbogens ergeben sich statistisch keinerlei Unterschiede zwischen den Testbedingungen.

Die farbige Benutzungsoberfläche wurde von allen BenutzerInnen unabhängig vom Zeitpunkt der Testung als eher "eindeutig", tendenziell "übersichtlicher" und weniger "vorgeschrieben" beurteilt (siehe Tabelle 17). Sobald also die Farbe zur unmittelbaren Handlungssteuerung eingesetzt wird, ergeben sich offensichtlich Vorteile für den Einsatz von Farbe bei der Gestaltung von Benutzungsoberflächen nicht nur im subjektiven Präferenzurteil, sondern auch im Performanzbereich.

2. Farbe und individuelle Systemanpassung

Um herauszufinden, wie die individuelle Farbauswahl sich auf die Benutzbarkeit auswirkt, wurden zwei empirische Untersuchungen durchgeführt: eine experimentelle Vergleichsstudie und eine elektronische Umfrage. Die Bedeutung des Konzeptes der Individualisierbarkeit wird von Ulich (1978, 1985, 1991) besonders hervorgehoben und wurde von Ackermann (1987) empirisch im Gestaltungsbereich der Dialogschnittstelle untersucht. Um das Kriterium der "Individualisierbarkeit" auch empirisch im Gestaltungsbereich der Ein-/Ausgabeschnittstelle bezüglich der Farbgebung untersuchen zu können, wählten wir eine Softwareumgebung aus, welche dem Benutzer diese Möglichkeit in ausreichendem Maß zur Verfügung stellt. Wir entschieden uns daher für die Windows 3.0-Umgebung, in der dem Benutzer vielfältige Farbeinstellungsmöglichkeiten angeboten werden.

2.1 Ergebnisse einer experimentellen Vergleichsstudie

Zunächst interessierte es uns, welche Farbeinstellungen von Benutzern gewählt werden, wenn die Ausgangssituation unterschiedlich ist: monochrome versus farbige Ausgangseinstellung. Wir stellen also die Hypothese auf, daß sich die individuellen Farbeinstellungen in Abhängigkeit von der Ausgangssituation ("monochrom" oder "color") systematisch hinsichtlich ihrer ergonomischen Qualität unterscheiden. Um diese Frage empirisch untersuchen zu können, wählten wir den weitverbreiteten Aufgabenkontext der Textverarbeitung und entwickelten ein Bewertungsraster für die von den Benutzern vorgenommenen Farbeinstellungen.

Im Aufgabenkontext von Textverarbeitung (Write™ unter Windows 3.0™) wurde Benutzern die Möglichkeit gegeben, sich die ihnen jeweils individuell zusagende Farbkonfiguration einstellen zu können (Haller et al. 1991). In der Ausgangssituation wurde eine monochrom ("schwarz-weiß") Vorgabe einer nach heutigen ergonomischen Erkenntnissen zulässigen Farbkonfiguration gegenübergestellt (siehe Bewertungsraster im Anhang).

Aus der Literatur (Kokoschka 1981; Murch 1985; van Nes et al. 1987; DIN 66 234 Teil 5 Beiblatt 2 1988) wurden zunächst alle bekannten Vor- und Nachteile von Farbkombinationen für verschiedene Typen von Bildelementen zu einem Bewertungsraster verdichtet (siehe Anhang). Zwei Arten von Belastungsquellen sind dabei zu berücksichtigen: physiologische Bedingungen und psychologische, bzw. kognitive Bedingungen. Es konnten 6 spezifische und 6 allgemeine Regeln zur Beurteilung von

Farbkonfigurationen abgeleitet werden, deren Bedingungen **nicht** erfüllt sein dürfen, um eine ergonomisch verträgliche Farbkonfiguration zuzulassen. Je nach Schweregrad des Verstoßes werden die einzelnen Regeln noch mit Gewichtungskoeffizienten versehen; ein schwerer Verstoß ist dadurch gekennzeichnet, daß dieser Aspekt in der Literatur besonders hervorgehoben wurde oder einfach öfters erwähnt wurde. Diese 12 Regeln legen in dem mehrdimensionalen Gestaltungsraum die "unzulässigen" Bereiche fest. Alle nicht von den Regeln erfassten Bereiche können zur Zeit als zulässige Konfigurationen gelten. Mit diesem Bewertungsraster kann nun die "software-ergonomische Güte" einer Farbeinstellung bestimmt werden.

In der experimentellen Untersuchung wurden 17 BenutzerInnen gebeten, pro Aufgabenblock einen vorgegebenen Text zu überarbeiten. Acht BenutzerInnen (2 Frauen, 6 Männer; Durchschnittsalter 29.6 ± 11.8 Jahre) begannen mit der schwarz-weiß Vorgabe (Gruppe-M). Die anderen 9 BenutzerInnen (6 Männer, 3 Frauen; Durchschnittsalter 29.0 ± 13.5 Jahre) fingen mit der Farbvorgabe an (Gruppe-C). Als abhängige Variable wurde die "software-ergonomische Güte" der individuellen Farbeinstellung nach Beendigung des Aufgabenblockes gemessen. Zum Schluß gaben die BenutzerInnen ihren subjektive Grad der Zufriedenheit mit der letzten Farbeinstellung auf einer kontinuierlichen Skala "Wie sind Sie mit Ihrer Auswahl zufrieden" von 0% bis 100% an.

Zur Kontrolle wurden folgende Variablen erhoben: Alter, Geschlecht, Farbschwäche (Gruppe-M: 0; Gruppe-C: 0), Vorerfahrung mit Computersystemen allgemein (Gruppe-M: 98 ± 176 std.; Gruppe-C: 59 ± 70 ; $p(t\text{-Test}) \leq .566$), mit Textverarbeitungsprogrammen (Gruppe-M: 4.0 ± 2.9 Jahre; Gruppe-C: 2.7 ± 2.2 Jahre; $p(t\text{-Test}) \leq .362$), mit farbigen Fenster-Oberflächen (Skalenbereich: 0="keine Erfahrung", 1="monochrom", 2="color"; Gruppe-M: 0.5 ± 0.8 ; Gruppe-C: 1.0 ± 1.2 ; $p(t\text{-Test}) \leq .335$) und mit sonstigen farbigen Oberflächen (Skalenbereich: 0="keine Erfahrung", 1="monochrom", 2="color"; Gruppe-M: 1.13 ± 0.99 ; Gruppe-C: 1.11 ± 1.27 ; $p(t\text{-Test}) \leq .980$), insbesondere mit der Windows 3.0-Umgebung (Skalenbereich: 0="keine Erfahrung", 1="monochrom", 2="color"; Gruppe-M: 0.88 ± 1.25 ; Gruppe-C: 0.22 ± 0.67 ; $p(t\text{-Test}) \leq .191$), Häufigkeit der Farbänderungen vor der Untersuchung "zu Hause" (Skalenrange: 0="nein", 1="einmal", 2="halbjährlich", 3="vierteljährlich", 4="monatlich", 5="immer wieder"; Gruppe-M: 0.5 ± 0.8 ; Gruppe-C: 1.7 ± 2.1 ; $p(t\text{-Test}) \leq .153$), sowie die Anzahl der getätigten Farbänderungen während des Aufgabenblockes (Gruppe-M: 6.9 ± 2.1 ; Gruppe-C: 4.6 ± 3.1 ; $p(t\text{-Test}) \leq .097$). Vor und nach der Textbearbeitung wurde als weitere Kontroll-Variable die subjektive Einschätzung des Augenzustandes mit einem Augenzustandsbogen erfaßt.

Während der Textbearbeitung wurden die Benutzer vom Testleiter angehalten, eine ihnen jeweils zugehörige Farbeinstellung vorzunehmen. Nach Ablauf von ca. 15 Minuten wurde die Aufgabenbearbeitung abgebrochen und die eingestellte Farbkonfiguration mit Hilfe der 12 Regeln des Bewertungsrasters beurteilt. Nur zwei BenutzerInnen der Color-Gruppe wurden mit Null-Fehlerpunkten bewertet, weil sie keine, bzw., nur eine unkritische Farbänderung vorgenommen hatten. Die Mono-Gruppe erreichte eine mittlere Fehlerpunktzahl von 4.1 (± 1.2 std.), die Color-Gruppe von 3.3 (± 3.0 std.); dieser Mittelwerts-

unterschied ist statistisch jedoch nicht bedeutsam (t -Wert=0.69, $df=15$, $p(t\text{-Test})\leq 0.484$)¹. Je mehr Einstellungsänderungen vorgenommen wurden, desto ergonomisch ungünstiger fiel die individuell eingestellte Farbkonfiguration aus ($R=0.583$, $p(R)\leq 0.023$). Der Mittelwertsunterschied in der Änderungsrate zwischen den beiden Gruppen ist statistisch nicht bedeutsam (s.o.); jedoch ist die Änderungsrate insgesamt über beide Gruppen hinweg abhängig von der Vorerfahrung mit farbigen Fensteroberflächen ($R= 0.526$, $p(R)\leq 0.044$), insbesondere mit der Windows-Umgebung ($R=0.566$, $p(R)\leq 0.028$), sowie allgemein mit anderen farbigen Oberflächen ($R=0.577$, $p(R)\leq 0.024$).

Hinsichtlich der subjektiven Einschätzung des Augenzustandes ergaben sich Unterschiede im Sinne einer Belastung weder zwischen den beiden Gruppen, noch zwischen Beginn und Ende der Aufgabebearbeitung. Zum Abschluß war die Gruppe-M mit $83.4\% \pm 17.2\%$ und die Gruppe-C mit $84.7\% \pm 12.2\%$ zufrieden mit der gewählten Farbeinstellung; dieser Mittelwertsunterschied ist nicht signifikant (t -Wert=-0.18, $df=15$, $p(t\text{-Test})\leq 0.859$); beide Gruppe sind also gleichermaßen mit ihrer jeweils erreichten Konfiguration zufrieden.

Unsere Hypothese konnten wir zwar nicht bestätigt finden, dennoch ist das gefundene Ergebnis von besonderem Interesse. Keiner der beiden Gruppen ist es gelungen, eine weitgehend fehlerfreie Farbeinstellung auszuwählen. Da die BenutzerInnen offenbar keinerlei Wissen über ergonomisch sinnvolle Farbkombinationen haben, ist es jetzt im nachhinein leicht erklärlich, daß sich die aktuelle Farbkonfiguration mit zunehmenden Farbeinstellungsänderungen von einer ergonomisch sinnvollen Einstellung entfernt. Es interessierte uns daher, herauszufinden, wieoft und in Abhängigkeit wovon Benutzer in der Praxis - sozusagen "vor Ort" - Farbeinstellungen vornehmen.

2.2 Ergebnisse einer Umfrage unter ETH-Mitarbeitern

Wir entschieden uns daher, eine Umfrage unter ETH-MitarbeiterInnen durchzuführen (Hegner et al. 1991). Um den Aufwand für die edv-mäßige Erfassung der Daten gering zu halten, wurde ein elektronischer Fragebogen entwickelt, welcher die Antworten gleich in einem für das vorgesehene statistische Auswertungsprogramm (Macintosh, SYSTATTM) geeigneten Datenformat anlieferte. Der elektronische Fragebogen konnte von allen ETH-MitarbeiterInnen, welche über einen email-Anschluß verfügten, aufgerufen und über eine VT-100 Terminalemulation beantwortet werden. Dieser elektronische Fragebogen war den ETH-MitarbeiterInnen für vier Wochen im Zeitraum vom 7.Juni bis 7.Juli 1991 frei zugänglich.

Es wurden insgesamt 202 elektronische Fragebögen ausgefüllt. Von diesen 202 Fragebögen wurden während der interaktiven Beantwortung 21 vorzeitig abgebrochen, sodaß noch 181 auswertbare Fragebögen übrig blieben. Die Rücklaufquote läßt sich nur schwer abschätzen, da nicht bekannt ist, wieviele der 3.940 wissenschaftlichen MitarbeiterInnen einen eigenen Rechneranschluß haben und von

¹ Um den rein quantitativ vorhandenen Unterschied zugunsten der Color-Gruppe auch gegebenenfalls inferenzstatistisch nachweisen zu können, wäre eine Stichprobengröße von 70 BenutzerInnen pro Gruppe notwendig (siehe Bortz 1984 S. 507).

einem der 169 ETH-Informatik-Koordinatoren über den interaktiven Fragebogen rechtzeitig informiert worden sind. Es existieren 21 Netzwerke, in denen 1394 angeschlossene Computer betreut werden. Wenn wir annehmen, daß die BenutzerInnen in Netzwerken am besten informiert worden sind und zu jedem Computer im Netzwerk genau ein Benutzer gehört, dann beträgt die Rücklaufquote 14%.

Der Fragebogen bestand aus 20 Fragekomplexen zu folgenden Themenbereichen: (a) Fragen zur Person, (b) Fragen zur EDV-Vorerfahrung (Anzahl Jahre, Kenntnisse verschiedener Computertypen [Sun-Workstations, IBM-PCs, Macintosh-PCs, Host-Terminals, sonstige] und Bildschirmmonitore [monochrom vs. color] , Nutzung verschiedener Softwareprogramme, etc.), (c) allgemeine Einschätzungen zur Verwendung von Farbe, (d) Gründe für bzw. gegen Farbmonitore, -software, etc., (e) individuelle Konfiguration von Farboberflächen.

Ca. 70% der ETH-MitarbeiterInnen würden sich bei freier Wahl für einen Farbbildschirm entscheiden, obwohl nur ca. 50% ihre Effizienz beim Arbeiten mit "color" Bildschirmen besser einschätzen als mit "monochrom". Auf die Frage: "Ist die Belastung bei color oder monochrom größer ?" antworteten 40% "gleich groß" und 30% "größer bei color".

Von allen Benutzern eines Color-Monitors [N=136] ändern 93% die vorgegebene Farbeinstellung!

Tabelle 20: Übersicht über die genannten Gründe für die Änderung der Farbeinstellung (N=136)

[Mehrfachantworten waren möglich]	N	(N/136)
"Vorgabe ist nicht nach meinem Geschmack"	92	68%
"aus Neugierde"	45	33%
"Kontrast ist zu schwach"	41	30%
"für eine individuell einzigartige Farbeinstellung"	14	10%

Personen, die aus Neugierde ihre Farbeinstellung ändern, tun dies signifikant *häufiger* als alle anderen Gruppen [$p(\text{Chi}^2\text{-Test}) \leq 0.037$]; ebenso ändern Personen, die eine einzigartige Farbeinstellung wünschen, fast signifikant *häufiger* als alle anderen Gruppen [$p(\text{Chi}^2\text{-Test}) \leq 0.058$]. Personen dagegen, die durch eine Änderung der Farbeinstellung den Kontrast verbessern wollen, tun dies signifikant *seltener* als alle anderen Gruppen [$p(\text{Chi}^2\text{-Test}) \leq 0.019$].

Benutzer von Macintosh-Farbbildschirmen (N=80) ändern die Farbeinstellungen signifikant *häufiger* als die Benutzer anderer Systeme [$p(\text{Chi}^2\text{-Test}) \leq 0.001$]. Umgekehrt ändern die Benutzer von IBM-PC Farbbildschirmsystemen (N=46) die Farbeinstellungen signifikant *seltener* als die Benutzer anderer Systeme [$p(\text{Chi}^2\text{-Test}) \leq 0.001$].

2.3 Konsequenzen für die individuelle Farbgestaltung

Aus den beiden zu letzt beschriebenen empirischen Untersuchungen geht klar hervor, daß (1.) ein hoher Prozentsatz der BenutzerInnen eine Änderung der voreingestellten Farbkonfiguration vornimmt und, daß (2.) diese BenutzerInnen wahrscheinlich sehr wenig bis gar kein Wissen über ergonomisch sinnvolle Farbkombinationen haben.

Um dieses Problem zu lösen, gibt es grundsätzlich zwei Möglichkeiten: (1.) die BenutzerInnen zu qualifizieren, und (2.) die Gestaltungsmöglichkeiten einzuschränken. Die Qualifizierung könnte über ein Tutorial, ein entsprechend gestaltetes Hilfesystem, über ein mit regelbasiertem Wissen ausgestattetes Konfigurierungsprogramm oder sonstige Qualifizierungsmaßnahmen (Schulung, etc.) erfolgen. Die zweite, eher pragmatische Möglichkeit besteht darin, dem Benutzer eine ausreichend große Menge an vorgegebenen Farbkonfigurationen ("configuration sets") zur Auswahl zur Verfügung zu stellen.

3. Fazit

Da nicht auszuschliessen ist, daß die Effekte von Farbe bei der Gestaltung von Benutzungsoberflächen offenbar in einer Größenordnung liegen, welche sich im Rahmen einer experimentellen Untersuchung mit einer Versuchsteilnehmerzahl von 10 bis 20 BenutzerInnen nicht in jedem Falle messen lassen, bleibt eine weitere Abklärung vorerst zukünftigen Untersuchungen vorbehalten. Man unterscheidet daher auch in der angewandten Statistik zwischen einem statistisch bedeutsamen und einem praktisch relevanten Unterschied. Im Rahmen der hier vorgestellten Untersuchungen können aufgrund der gewählten Stichprobengrößen nur Unterschiede mit einer Effektgröße von 0.5 und darüber als praktisch relevant nachgewiesen werden (siehe hierzu Bortz 1984 S.508ff). In diesem Sinne konnte nur bei einem von drei Experimenten ein praktisch relevanter Unterschied im Performanzbereich zugunsten von Farbe gefunden werden.

Interpretiert man die subjektiven Beurteilungen als sensitive Meßgrößen, so kann man aus den gefundenen Ergebnissen eine Überlegenheit von farbigen Benutzungsoberflächen ableiten. Im Performanzbereich läßt sich eine Überlegenheit durch Farbeinsatz dann nachweisen, wenn die Farbgebung handlungsleitend zur Erhöhung der Transparenz der Dialogstruktur eingesetzt wird.

Für die individuelle Konfigurierbarkeit von farbigen Oberflächen konnte insofern ein wichtiger Befund erhoben werden, als offenbar keiner der VersuchsteilnehmerInnen ein hinreichendes Wissen über eine sinnvolle Farbgebung besaß. Um diesem Umstand auf eine einfache Weise Rechnung zu tragen, könnte der Gestaltungsspielraum auf eine Menge von vorgegebenen, ergonomisch vertretbaren Einstellungen beschränkt bleiben. Viele Farbbildschirme lassen die ergonomisch erforderliche Kontrastgüte vermissen, sodaß die Benutzer versuchen, über eine Änderung der Farbgebung den Kontrast zu erhöhen.

4. Literatur

- Ackermann, D.: Handlungsspielraum, mentale Repräsentation und Handlungsregulation am Beispiel der Mensch-Computer-Interaktion. Dissertation an der Universität Bern. Eidgenössische Technische Hochschule, Institut für Arbeitspsychologie, Zürich 1987
- Apenburg, E.: Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung. Psychologie und Praxis, Zeitschrift für Arbeits- und Organisationspsychologie, 30 (N.F.4) (1986) 1, S. 3-14
- Baumgartner, M., Bauer, D., Messmer, B.: Erhöhung der Transparenz durch farbige Benutzungsoberflächen bei Simulationsprogrammen. unveröffentlichte Gruppensemesterarbeit. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1991
- Bortz, J.: Lehrbuch der empirischen Forschung. Springer, Berlin Heidelberg New York 1984
- Christ, R.D.: Review and analysis of color coding research for visual displays. Human Factors, 17 (1975) 6, pp.542-570
- Cotting, S., Kohli, D.: Erhöhung der Transparenz durch farbige Benutzungsoberflächen bei Hypertextsystemen. unveröffentlichte Gruppensemesterarbeit. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich, 1991
- DIN 66 234, Teil 5: Bildschirmarbeitsplätze - Codierung von Information. Beiblatt 2 "Farbkombinationen", 1988
- Haller, S., Perjes, B., Stoll, D., Züger, P.: Farbgestaltung von Bildschirmoberflächen: Einfluß von unterschiedlichen Startbedingungen auf die Güte der individuellen Farbauswahl. unveröffentlichte Gruppensemesterarbeit. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1991
- Hegner, D., Hüppin, R., Kälin, A., Lerch, S.: Nutzen und Benutzbarkeit von Farbe bei Computeranwendungen. unveröffentlichte Gruppensemesterarbeit. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1991
- Kokoschka, S.: Visuelle Leistung am Bildschirmarbeitsplatz in Abhängigkeit von der Zeichenfarbe. Die Farbe, 29 (1981) 1, S. 8-16
- Loeliger, T., Uster, M., Spiniello, C., Salvatore, A. Erhöhung der Transparenz durch farbige Benutzungsoberflächen bei Spielprogrammen. unveröffentlichte Gruppensemesterarbeit. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1991
- Lynch, P.J.: The surface anatomy of birds, version 1.2. Biomedical Communications - Yale Center for Advanced Instructional Media, Yale University School of Medicine USA 1988
- McDonald, J.E., Molander, M.E., Noel, R.W.: Color-coding categories in menus. In: Soloway, E., Frye, D., Sheppard, S.B. (eds.): Proceedings of the Conference CHI'88 "Human Factors in Computing Systems", Washington, ACM Press (1988), pp.101-106
- Munns, M.: Some effects of display symbol variation upon operator performance in aircraft interception. Perceptual and Motor Skills, 26 (1968), S. 1215-1221
- Murch, G.M.: Colour Graphics - Blessing or Ballyhoo? Computer Graphics Forum, 4 (1985), pp.127-135
- Pazhitnov, A.: Tetris. Spectrum HoloByte Inc., 2061 Challenger Dr. Alameda CA. 94501, USA 1988
- Rauterberg, M.: Benutzungsorientierte Benchmark-Tests: eine Methode zur Benutzerbeteiligung bei der Entwicklung von Standardsoftware. In: Spinass, P., Rauterberg, M., Strohm, O., Waerber, D., Ulich, E. (Hrsg.): BOSS-Projektbericht Nr. 6, Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1991

- Spinas, P.: Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen. Dissertation an der Universität Bern. Institut für Arbeitspsychologie, Eidgenössische Technische Hochschule, Zürich 1987
- Ulich, E.: Über das Prinzip der differentiellen Arbeitsgestaltung. Industrielle Organisation, 47 (1978), S. 566-568
- Ulich, E.: Einige Anmerkungen zur Software-Psychologie. Sysdata, 10 (1985), S. 53-55
- Ulich, E.: Arbeitspsychologie. Poeschel, Stuttgart 1991
- van Nes, F.L., Juola, J.F., Moonen, R.J.A.M.: Attraction and distraction by text colours on displays. In: Bullinger, H-J., Shackel, B. (eds.): Proceedings of the Conference INTERACT'87 "Human-Computer Interaction", North-Holland, Amsterdam 1987, pp. 625-630
- Wright, W.: SimCity - Gebrauchsanweisung für PC und Kompatible. Infogrames 1989

5. Anhang

Bewertungsraster für Farbkombinationen bei Windows 3.0

Spezifische Regeln für die zu bewertenden Bildelemente, bzw. -kombinationen:

Regel 1: Bewertungsraster für Text/Hintergrundkombination.

[Bewertungspunkte: 0 = keine negativen Auswirkungen, 1 = schlechte Kombination
Kombination]

Text- farbe	Hintergrundfarbe							
	schwarz	weiss	purpur	blau	cyan	grün	gelb	rot
schwarz	3	0	0	1	0	0	0	1
weiss	0	3	0	0	1	1	1	0
purpur	0	0	3	1	1	3	1	1
blau	1	0	1	3	0	1	0	3
cyan	0	1	1	0	3	1	1	3
grün	0	1	3	1	1	3	1	3
gelb	0	1	1	0	1	1	3	0
rot	1	0	1	3	3	3	0	3

Regel 2: Reines Blau wurde für Text, dünne Linien und feine Bildelemente eingesetzt.

Regel 3: Es gibt nebeneinanderliegenden Farben, die sich nur im Blauanteil unterscheiden.

Regel 4: Bildelemente, die eine sofortige Aufmerksamkeit hervorrufen sollen, sind weder durch starke Sättigung, noch durch große Helligkeit hervorgehoben.

Regel 5: Bildelemente, die primär für Aktionen oder Aufforderung zur Reaktion, bzw. Eingabe vorgesehen sind, sind nicht durch warme Farben dargestellt. Bildelemente, die für die Ausgabe von Zuständen oder Hintergrundinformationen vorgesehen sind, sind nicht durch kalte Farben dargestellt.

Regel 6: Nebeneinanderliegende Flächen unterscheiden sich weder in ihrer Farbe, noch in ihrer Helligkeit. Eine achromatische Trennlinie ist auch nicht vorhanden.

Die jeweils zutreffende Regel 1 bis 6 ist in der folgenden Tabelle mit dem entsprechenden Gewichtungsfaktor zu multiplizieren und der sich ergebende Wert als Fehlerpunkt einzutragen.

Regel Nr.

[in der Tabelle stehen Gewichtungsfaktoren]

Bildelemente	1	2	3	4	5	6	Fehlerpunkte
Text / Fensterhintergrund	2	1	0	0	0	0	1)
Text / Menüleiste	1	1	0	0	0	0	2)
Text / aktive Titelleiste	1	1	0	0	0	0	3)
Text / inaktive Titelleiste	1	1	0	0	0	0	4)
Desktop-Fläche	0	0	1	1	1	1	5)
Programmarbeitsbereich	0	0	1	1	1	1	6)
Fensterhintergrund	0	0	1	1	1	1	7)
Menüleiste	0	0	1	1	1	1	8)
aktive Titelleiste	0	0	1	1	0	1	9)
inaktive Titelleiste	0	0	1	1	1	1	10)
aktiver Rahmen	0	0	1	1	0	1	11)
inaktiver Rahmen	0	0	1	1	1	1	12)
Fensterrahmen	0	1	1	1	1	1	13)

Bildlaufleisten	0	0	1	0	0	1	14)
-----------------	---	---	---	---	---	---	-----

Allgemeine Regeln für den gesamten Bildschirmaufbau:

- Regel 7: Bei einfachen Bildern werden keine Komplementärfarben verwendet (z.B. rot mit gelb, grün mit blau).
- Regel 8: Bei großflächigen Symbolen, bzw. dicklinigen Zeichen sind die Kombinationen blau/schwarz, weiss/gelb, grün/cyan, rot/purpur verwendet worden; bei dünnlinigen Zeichen kommen Kombinationen mit weiss/..., gelb/..., grün/... und cyan/... vor.
- Regel 9: Es werden gleichzeitig stark gesättigte, spektral extreme Farben nebeneinander verwendet (z.B. cyan, bzw. blau mit rot, gelb mit purpur, etc.).
- Regel 10: Es werden viele verschiedene Farben (mehr als 6) eingesetzt [als Ausnahme können eventuell alle nach ihrem Spektrum geordneten Farben gelten].
- Regel 11: Grüne oder rote Farben werden in der Peripherie des Bildschirms verwendet.
- Regel 12: Zusammengehörnde Bildteile sind mit unterschiedlichen, bzw. nicht zusammengehörnde Bildteile mit gleichen oder sehr ähnlichen Farben codiert.

[Gewichtung: 1=geringer Verstoß, 3=schwerer Verstoß]	Anzahl Regelanwendungen	Gewichtung	Fehlerpunkte
Regel 7		* 1	15)
Regel 8		* 1	16)
Regel 9		* 3	17)
Regel 10		* 3	18)
Regel 11		* 1	19)
Regel 12		* 1	20)

Zur Berechnung der Fehlerpunktzahl insgesamt sind die einzelnen Fehlerpunkte 1) bis 20) aufzuaddieren. Je kleiner dieser Wert ist, desto besser ist die Farbgebung der jeweiligen Oberfläche. Im folgenden werden die 17 empirisch gefundenen Farbeinstellungen (StartMono: colconfP1 .. P8; StartColor: colconP9 ... P17) aus der Vergleichsstudie mit Write unter Windows 3.0 aufgelistet.

INHALT der CONTROL.INI Datei:

```
[current]
color schemes=StartColor

[color schemes]
StartColor=FF8080,E8FFFF,FFFFFF,0,8080,0,808000,C0C0C0,FFFFFF,80FFFF,808080,0,C0C0C0
StartMono=FFFFFF,FFFFFF,FFFFFF,0,FFFFFF,0,0,FFFFFF,FFFFFF,0,FFFFFF,0,FFFFFF
colconfP1=808040,FFFFFF,FFFFFF,0,808080,800000,828282,FFFF80,800000,FFFF80,828282,0,8
colconfP2=400040,FFFF80,FFFFFF,0,800080,0,FF80FF,FF80FF,0,FFFF80,FF8080,0,FFFFFF
colconfP3=8000,FFFFFF,8000,FFFFFF,C0C0C0,0,828282,FFFFFF,FFFFFF,0,800040,0,FFFFFF
colconfP4=C0C0C0,808040,C0C0C0,0,808080,0,C0C0C0,808040,0,FFFFFF,828282,0,808040
colconfP5=FFFF80,FFFFFF,FF00FF,0,FFFFFF,0,0,FFFF00,FFFFFF,80FF00,FFFFFF,0,FFFFFF
colconfP6=828282,E8FFFF,FFFFFF,8000,FFFFFF,80,8000,FFFFFF,FFFFFF,4000,FFFFFF,8080,FF
colconfP7=FFFFFF,FF,FF00,0,FFFF,0,FFFF80,FFFFFF,FF0000,0,FFFFFF,0,FFFFFF
colconfP8=FFFFFF,E8FFFF,FFFFFF,80,808080,80,808080,408080,FFFFFF,408080,FFFFFF,0,FFF1
colconfP9=FF8080,E8FFFF,FFFFFF,0,FF00FF,0,808000,C0C0C0,FFFFFF,400000,FF80FF,0,C0C0C0
colconfP10=40,828282,FFFFFF,0,FFFFFF,0,8080FF,C0C0C0,0,80,808080,0,C0C0C0
colconfP11=FF8080,E8FFFF,FFFFFF,0,8080,0,800040,C0C0C0,FFFFFF,80FFFF,808080,0,C0C0C0
colconfP12=FF8080,E8FFFF,FFFFFF,0,8080,0,808000,C0C0C0,FFFFFF,80FFFF,808080,0,C0C0C0
colconfP13=FF8080,80FFFF,0,FFFFFF,FFFFFF,0,8080FF,C0C0C0,0,80FFFF,0,0,80FF
colconfP14=0,E8FFFF,FFFFFF,0,FF00FF,FFFFFF,FFFF80,C0C0C0,0,80FFFF,808080,0,C0C0C0
colconfP15=80FF,8080FF,FFFF00,0,808080,0,808000,C0C0C0,FFFFFF,80FFFF,808080,0,C0C0C0
```

colconfp16=FF8080,FFFFFF,FFFFFF,0,8080,0,A00000,FFFF,FFFFFF,80FFFF,808080,0,C0C0C0
colconfP17=400080,408080,FF0000,FFFF,C0C0C0,0,FF0080,FFFF00,FFFFFF,8000FF,FF0000,0,C0