Stimulus Delivery on the Web: What Can Be Presented when Calibration isn't Possible

John H. Krantz[1]

At the heart of most psychological experiments is the stimulus. It is the feature of an experiment over which the researcher has most control. However, as a concept, the stimulus has generally received very little attention for such a central feature of our experiments. There are some notable exceptions (Gibson, 1960), but generally, it appears that psychology has a tendency to assume that the stimulus is sufficiently robust that any error attributable to the stimulus or variations in the stimulus does not affect the outcome of the experiment. Such a position is contradicted by the relatively small but vital literature on the need to calibrate video terminals and other methods of stimulus presentation (e.g., Brainard, 1998; Krantz, 2000; Metha, Vingrys & Badcock, 1993; Olds, Cowan, & Jolicoeur, 1999).

There are many circumstances that require careful calibration so that the stimulus does not become a confound in an experiment. Let me give one simple example that will be amplified below. One common assumption is that the screen is the same over the entire surface. That is the stimulus will be identical regardless of the position on the screen. The luminance, the size, the color, all of the physical attributes of the stimulus are identical regardless of position. However, the luminance is not identical across the screen surface (Cook, Sample, & Weinreb, 1993; Hu & Klein, 1994). Krantz (2000) found a variation in the luminance of over 20% on a CRT and over 40% on a LCD. This variation will affect the brightness and contrast of the stimulus depending on location. Both of these variations can affect the behavioral responses of the participants (Krantz, Silverstein, & Yeh, 1992; Snyder, 1988). This need for careful monitor calibration of our stimulus presentation device is in direct conflict with the diffuse nature of the Web. Calibration requires knowing what hardware one is using. With the Web, one cannot be certain what type of equipment the person is using at the other end of the connection. While some information such as browser, operating system and even resolution and number of colors can be

[1] Hanover College, USA. E-mail: krantzj@hanover.edu
I would like to thank Barrie Woods for his great help as a reviewer of this chapter. Beyond that, I would like to thank him for years of friendship and intellectual companionship, particularly in our playful moments.

gathered, much needed knowledge about the monitor is still missing. What one can get over the Web is the command from the computer, i.e., the input to the display. Proper calibration requires the ability to measure the output of the device against the input and knowledge of the position of the user and even the lighting in the room. With the Web, one may be half a world away from that device making this type of precise measurement a moot possibility. Calibration requires that the device is always in the same stable state for each participant. With the Web, one cannot know the state of the device, how the user might have adjusted it, or be assured that it has been properly warmed up. I could go on, but I think I have made the point that there is no way, with current technology, to do adequate calibration for a device on the Web. This situation does not lead to the conclusion that Web research cannot be adequately or successfully be done. Several studies have been able to use the Web quite successfully to present stimuli and collect responses (e.g., Krantz, Ballard, & Scher, 1997; Krantz & Dalal, 2000; Reips, 1995). Thus, research is possible even when we cannot calibrate.

The conclusion that I am going to promote in this chapter is that researchers must be careful about how a stimulus is constructed. Towards this end, I will review the issues regarding calibration and make suggestions on what can be done to avoid these problems. Given the brevity of this chapter, I will only discuss visual stimuli. However, there are many parallel issues for auditory stimuli that can make or break an online study (Welch & Krantz, 1996).

The Computer Monitor

The monitor attached to the computer will usually be one of three types of display: the cathode-ray tube (CRT), liquid-crystal diode, or plasma display. In each case, the display technologies take advantage of the way our visual systems functions as they produce the image. I will discuss three main features of our visual system that modern display technologies use to great advantage. They are the spatial summation, the temporal summation, and the trichromatic nature of our visual system. Let me briefly discuss each of these in turn.

Spatial Summation

The visual world is, at our scale of experience, continuous. Displays are not. They are made of discrete points called pixels. Displays may have resolutions such as 800 x 600 or 1024 x 768 or even higher on many newer displays. This resolution refers to the number of dots on the screen that are capable of producing the full range of colors of the display. There are usually more dots than that on the screen surface (Krantz, 2000, Snyder, 1988). These pixels and the individual dots that make them up are not continuous but have distinct edges around them. If you look through a magnifying glass onto the surface of your monitor you will notice these lines, often appearing black, around each dot on the screen surface. Yet, when we look at most monitor screens at a normal viewing distance, we see a continuous image and not a speckled surface. This is an example of spatial summation. The eye is very different from a camera. In a camera, and ideally on a monitor surface, each dot is independent of every other dot. Thus, if one pixel is white, the next pixel could be black, gray or white and it would not change the light output at that first dot. The mammalian eye does not work that way at all. By virtue of the receptive fields of the visual system, we sum the input across small regions of the retina (Kuffler, 1953). As a result of this fact, if two dots are placed close enough together we will see them as one. Under optimal conditions of illumination and contrast, the minimum visual angle for resolving two dots from one is about .5 to 1 arcmin (e.g. Morgan, 1991). So monitors are designed so that the dots on the screen are close enough together to make sure that they fall within the range for spatial summation. To make use of human spatial summation, monitor developers need to make some assumptions about the normal viewing distance of the user. With current technology, they can easily surpass this summation limit for most normal circumstances and for a wide range of stimuli.

Temporal Summation

Similarly, while stimulation in the natural world is continuous in time, displays take advantage of our temporal summation to present discontinuous stimulation that appears continuous in time. Just as moving two dots closer together on a surface will cause them eventually to appear as one, we can present two flashes closer together in time and eventually they will appear as one flash (in the fovea at a flicker rate of about 60 Hz). If they are spatially separated, they will appear to move from one location to the other which is called beta motion and occurs already at an alternation rate of about 10 Hz (Wertheimer, 1912). Thus, the frame rate and update rate on monitors are critical, both for smooth motion and, more importantly, so that we do not see the surface flicker. American television sets update the screen at 60 Hz, which is adequate for most circumstances. However, many computer monitors flicker at a higher rate, which is better for minimizing our ability to perceive the screen flicker (Bridgeman, 1998). These temporal responses of our eyes are the results of the temporal summation on neurons and the dynamics of cells such as the magnocellular cells of the lateral geniculate nucleus (Lennie, Trevarthen, Van Essen, & Waessle, 1990).

Color

The final characteristic that monitors take advantage of is the essentially trichromatic nature of our eye (Davidoff, 1977; Mollon, Cavonius, & Srenner, 1998). While we can perceive an enormous range of colors, we do so through a bottleneck of sorts, through the normal eye's three different classes of cones. Commonly, these three cones are called the red, green, and blue cones but that match is only approximate. Thus, while the world is full of a wide range of combinations of wavelengths, the monitor only needs to match the activity of these three classes of cones in our eyes to reproduce to a color in the natural world. So an enormous range of colors gets reduced to three separate values for the three primaries we have been taught since our earliest years of school. If the cones respond identically to two patches of light, there is no way for the brain to be able to tell if the same or different combinations of wavelengths generated these two patches. This is the basis of color matching. Thus, the fact that we are trichromatic makes it possible for paintings and color monitors to be in color. For the number of primaries needed for color reproduction is tied to the number of cones that we have. Because we have only three cones, we only need three primaries which makes it relatively easy to get the primaries close enough together to fit in our spatial summation limits as was discussed above. The formalization of color perception for the color normal person has been done in the CIE system (Silverstein and Merrifield, 1985). This system allows engineers to develop precise color matching systems such as color monitors.

Calibration of the Monitor

With this very basic introduction to monitors, we can now proceed to the issue of why calibration is needed. Given the only rare report in the literature about the calibration of the experimental monitor in experiments, even in the laboratory, it seems that the need for calibration should be justified. Basically, the principle reasons for calibration are that displays are not all identical, they are not perfectly stable, and they are not simple.

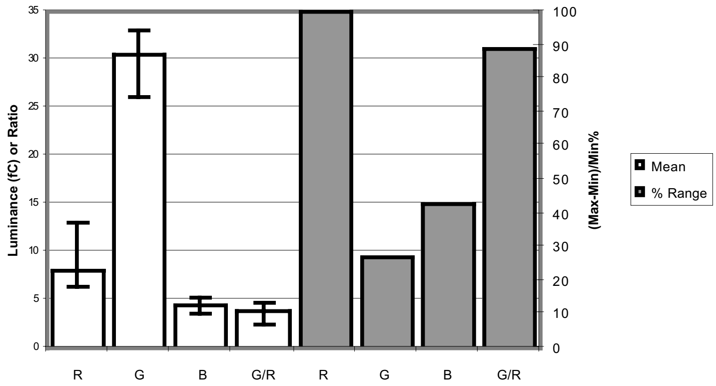

Figure 1. The light bars show the average luminance for the three primaries and the ratio of the green luminance to the red luminance (G/R) for six identical monitors. The monitors all have the same settings on them and have been warmed up for the same period of time. The error bars show the range of values for six monitors. The right bars (gray) converts the range into a percentage relative to the lowest value obtained across these six monitors. Note the inconsistency of these supposedly identical monitors set up to be identical.

Figure 1 shows the results of a very simple experiment. I took six identical monitors, bought at the same time and less than two years old at the time of this test. I set them all to the same settings of brightness, contrast, and color temperature. I warmed all of them up for the same period of time (Olds et al., 1999). Then I measured the luminance at maximum for each of the three primaries. In Figure 1, I plotted the mean and the range of these luminance values across this small sample of monitors. In addition, I converted the range into a percentage of the lowest value obtained for that primary, (Max-Min)/Min*100. Notice the wide variation in all of these values. The red gun varies by 100% across the six monitors. The green gun is the most stable in relative terms varying only 27% and that is still an enormous error. More importantly, notice the variation in the ratios of red to green gun luminance that is also shown. While the differences in absolute values can be important, the difference in the ratio is more critical. These ratios will influence the colors output on each of these monitors. In this sample the range of variation of the ratio of the red to green luminance is almost 90%. Thus, one cannot be certain to have the same color on these six monitors even given the same pixel values. This is just one example of an important way that monitors differ from each other.

In Figure 2, the same measurements are shown for a single monitor (one of the monitors from above). In this experiment, the monitor was measured as soon as turned on, then again after 25 minutes. In addition several settings of brightness and contrast were used ranging from the 50% setting to the 100% setting. As can be seen from this simple plot, even a given monitor changes from time to time. The intensity values vary greatly as is to be expected but notice the variation of the red to green gun. It varies by

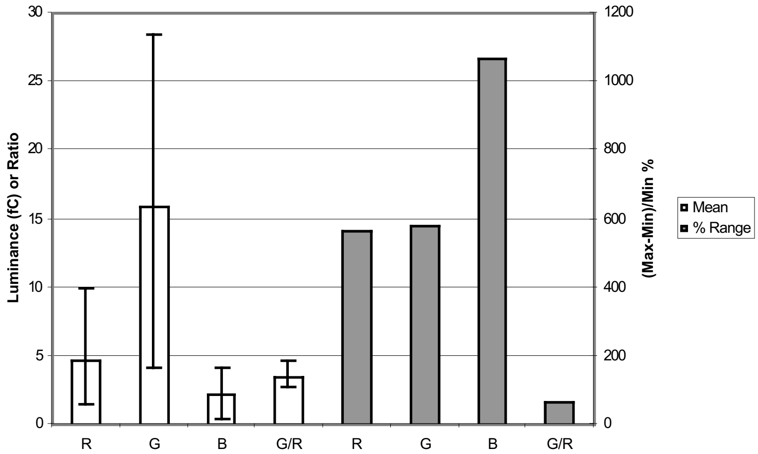

Figure 2. This figure shows how various settings on a monitor can affect its output. Conditions manipulated are time after warm up and the brightness and contrast setting on the display. Notice that these setting produced an even greater range of variation for the intensities than was seen across the monitors in Figure 1 when all of the monitors were at the same settings. The variation in the ratios for the green to red gun is about the same within as between the monitors.

over 60% across these settings or nearly as much as was seen across the six monitors in Figure 1. Thus, calibration is only good for one set-up on the monitor. In addition, calibration must be done regularly as the monitor will change over time. Finally, one should warm up the monitor, as that is when it is most stable. Most studies recommendations are for monitors to be warmed up for 30 minutes before calibration and the experiment (Metha et al., 1993; Olds et al., 1999).

The next calibration related issue deals with how the monitor both creates the images and how these functions interact with our visual system. A fuller treatment of the subject is found in Krantz (2000). For this chapter, let us briefly look at pixel interactions and the gamma of a monitor as representative issues.

Ideally each pixel should behave independently of every other pixel. This situation is what is required for a probability sample, and the monitor is a sampling device (Krantz, 2000). As an example, examine the figure found at: http://psychlab1.hanover.edu/SCiP/1999/CRTtestimage1anim.gif. This image is a simple alternating vertical and horizontal grating as this is possible for a monitor to draw. In each case the black lines are as dark as the screen can be and the white lines are as bright as possible. The screen is half black and half white and both gratings should appear at the same brightness if viewed far enough away for the lines to be spatially summed. If using a CRT, undoubtedly, the horizontal grating is brighter than the vertical grating. Measurements made by Krantz (2000) on a random selection of monitors found that in one case the vertical grating had only 20% of the mean luminance of the horizontal grating. In all but one of the seven monitors, the vertical grating had less that 60% of the mean luminance of the horizontal grating. This difference in the mean luminance of the vertical and horizontal grating is a result of the fact that on CRT's, the monitor cannot go from all the way off to all the way on and back again pixel by pixel. A LCD or plasma display may not illustrate this effect. These are digital displays and their pixels are more independent. Krantz (2000) found on one LCD that the vertical and horizontal grating had nearly the same mean luminance, which is what is desired.

The other issue is what is called the gamma of the display. To turn on a pixel on most computer monitors, one must use three values usually ranging from 0 to 255, one for each of the primaries. The range of values, not necessarily of luminance, is greater on some monitors. Ideally, a value of 200 on one primary would provide twice as much light output as a value of 100 on the same primary. There is not a nice linear relationship between the luminance output and the digital value for the pixel. The relationship is usually described by an exponential function called the gamma function (Davidoff, 1977, Olds et al., 1999). In short, without determining the gamma function for each of a monitor's primaries (and yes, each primary will have a different gamma) one cannot know either the luminance or, more importantly, the contrast for the stimuli on any given monitor. Given both the visual systems ability to adapt and the organization of receptive fields, contrast is the more important measure to know (Cornsweet, 1970).

So What Can Be Done?

Hopefully by this time the reader is fully aware of the complexity of this device called the monitor and perhaps a bit nervous about the presentation of stimuli on them. However, we do not need to go so far as to conclude that the situation is hopeless and that Web-based research is impossible unless we restrict such research to surveys. Yes, I would like psychological research to be much more careful about calibration and report much more about their calibration procedures than is common in most psychological reports at this time (Krantz, 2000). However, just as monitors take advantage of the visual system to successfully reduce the information transmitted and, thus, make their technology possible, psychological researchers can do the same and minimize the chance of falling afoul of monitor induced confounds. In addition, there are ways to anticipate some of the limits of monitors and we can develop stimuli to step wide of these limits. In this section of the chapter, I will give some practical hints on how to develop appropriate Web-based stimuli.

Taking Advantage of the Visual System

The first feature of the visual system that can be taken advantage of is its adaptive nature. There are many forms of adaptation from dark/light adaptation that helps adjust to the prevailing light level to chromatic adaptation that plays a role in color constancy. These mechanisms serve, collectively, to emphasize features of the visual world where changes occur, change in luminance, change in color, etc, and to minimize our normal sensitivity to areas where there are no changes. In other words, the average changes in luminance between monitors may not be noticeable because we will adapt to them. So we need not worry about the average differences between monitors but we will need to worry about contrast and color balance, which is exactly why the variation in the ratios of the red and green gun illustrated above are so problematic (Figures 1 and 2).

The contrast of the image plays an important role in many behavioral responses such as reaction time, accuracy, and legibility (e.g., Krantz et al., 1992). This measure is crucial and can alter results in studies using such dependent measures as reaction time. Even a participant's responses on a survey can be affected by legibility and fatigue from difficult reading. However, there seems to be an asymptotic value for the effect of contrast on behavioral responses (Krantz et al., 1992; Sivak, Olson, & Pastalan, 1981; Snyder, 1988). In other words, beyond a certain point, increasing the contrast will no longer affect behavioral responses. This asymptotic value for contrast is found with a contrast ratio of about 3:1 (Krantz et al., 1992; Sivak, et al., 1981, Snyder, 1988). This contrast ratio is the same as a Michaelson contrast of .5. If the behavior is robust and fine discriminations are not needed, a 2:1 contrast ratio is adequate (Penn & Silverstein, 1998). This value is relatively easy to achieve on most monitors. For instance, considering white as the stimulus, contrast ratios on the surface can be over 30: 1. The ratios are far smaller for colored stimuli. So reaching this minimum contrast needed seems rather easy as least for non-colored stimuli.

One factor that can ruin a careful stimulus is overhead lighting and light streaming in from the outside through a window. Some of the light will reflect off of the surface of the monitor and reduce the stimulus contrast on the monitor (Krantz, 2000; Krantz et al., 1992; Silverstein & Merrifield, 1985). The sun can easily wash out a screen, especially if a CRT is involved. Researchers should include instructions for the monitor to be in a room with shades drawn or at least not facing an open window as part of the instructions and informed consent for the Web-based study. Overhead lights are probably a necessary evil to be taken into account in the design of stimuli. Thus, aim for maximum contrast or very great contrast so that stimuli do not creep near the 3:1 asymptote. If the light levels are absolutely critical to the design for whatever reason, contrast or flicker or other variables that can be sensitive to light level, then the study is best limited to the laboratory.

Color presents a whole new range of problems for the researcher. Color can be used as an independent variable for at least two types of research purposes. First, color may be manipulated as an independent variable in order to understand our color perception better. Second, color may be manipulated because it is a powerful variable that can affect other behaviors. For example, take the use of color in visual search research. If the target is of a distinctly different color than the background then the target will seem to "pop-out" against the background (D'Zmura, 1991; Treisman & Gelade, 1980). Search is made much easier by this simple color manipulation. The research that seeks to understand our color perception is at such an advanced state that only the most carefully calibrated stimuli can be used (Wyszecki & Stiles, 1967). Thus, this research is not presently possible over the Web and will not be further discussed here. The rest of the comments refer to this second use of color as an independent variable.

While the requirements for color calibration are not as great in a search experiment as in a color matching experiment, the data presented earlier in

Figures 1 and 2 indicate that the color of stimuli may still vary across monitors. These are all the same type of CRTs. When sunlight, overhead lights and different technologies are considered, the situation will become worse. In many cases, it might not be a problem; just as there are adaptation mechanisms leading to constancy of brightness perception, there are mechanisms leading to constancy in color perception (Zeki, 1991). Thus, in many cases the perception of the color will not change even though the stimulus will change greatly provided the relative colors in the surround stay the same ratios.

Despite these constancies, any use of color that is critical and requires sure knowledge of the color to be displayed in the experiment is to be avoided. Color simply is too poorly controlled from monitor to monitor to be a useful independent variable in any but the most robust cases (Laugwitz, this volume). In addition, there is a sizeable percentage of the population that does not have normal color vision with many more men suffering from a color vision deficiency than women (Mollon, 1992). Their color perception differs significantly from the majority. In the laboratory, these people are usually screened and eliminated or treated as separate groups depending upon the nature of their color deficiency.

Still, color can be used in some cases as an experimental independent variable. One exception would be the case where the presence or absence of color is the independent variable. Here color is operating as a code, as in the visual search studies mentioned above (see also Laugwitz, this volume). Once could also vary a very small set of code colors in an experiment. Before doing so, the color should be pilot tested on several monitors and in several lighting conditions before using them on the Web. It is advisable to discuss how one validated the color independent variable in the method section.

In summary, researchers should avoid stimuli that require fine distinctions, as it is in this area that the lack of calibration will defeat an experiment. With regard to contrast, researchers should use very high contrast stimuli, avoid complex backgrounds that can reduce contrast, and ask users to reduce extraneous light as much as possible. With regard to color, this variable is not viable as an independent variable at the present time. There are simply too many ways that the output is unpredictable. One cannot tell on any monitor what is being seen on another monitor with sufficient reliability to be able to use it. It should be alright to use color as a variable against the lack of color however.

Avoiding Monitor Problems

Above, I discussed such monitor difficulties as lack of pixel independence and the non-linear gamma functions for the luminance of the primaries. There are other such problems just as the jaggies at the end of lines that can make letters less easy to read. The way to minimize the pixel independence problem while maintaining a high contrast image is not to use a real fine image. If the rate of change between the white and the black regions of an image is slower, then there will be less of a problem of losing contrast and luminance in horizontal lines.

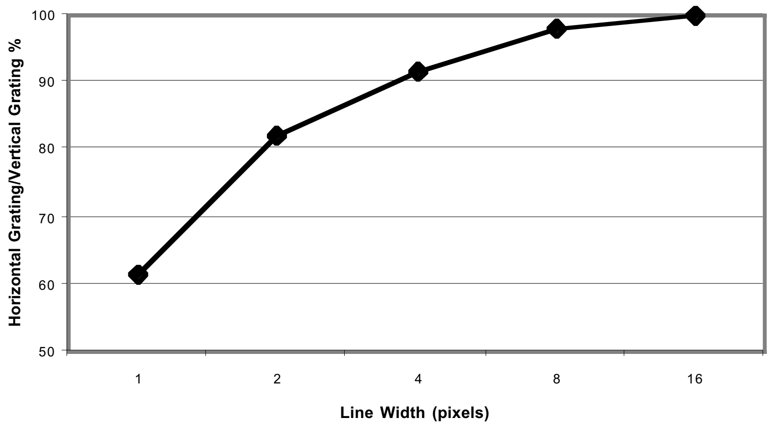

Figure 3. This figure plots the mean luminance of a horizontal grating relative to a vertical grating for lines of different widths from one monitor. A value of 100% indicates that the two gratings have the same mean luminance. Notice how the ratio rapidly approaches 100%. By the time the lines are 4 pixels wide, the two gratings are within 10% of each other.

Figure 3 shows some data from a warmed-up monitor for horizontal versus vertical gratings. The mean luminance of the vertical grating is divided by the mean luminance of the horizontal grating. If the pixels are completely independent, the ratio should be 100%. This ratio starts out far too small (62%) but is close to 100 when the lines are 4 pixels wide. These lines gratings are called square-wave gratings. The screen elements are either full on or full off. This problem can be reduced even for the 4 pixel wide lines using antialiasing (see below).

To avoid the problems of gamma, one should stick to very simple images with few levels of luminance as it is impossible to predict contract or luminance from a remote location. Knowing what is commanded to the monitor is simply not enough. Therefore, the simpler the image using as few luminance levels as possible, the more faithful the reproduction on a wide range of monitors and the more adequate the stimulus. This approach was taken by Krantz et al. (1997) in their replication of Fallon and Rozins's (1985) study and Wiggins, Wiggins, and Congers's (1968) study of determinants of female attractiveness. Senior, Phillips, Barnes, and David (1999) used a similar approach in their replication of Keating, Mazur, and Segall's (1977) study of the facial features associated with dominance. In both of these Web studies, the graphical stimuli were simple line drawings of black lines on a white background. These stimuli were of high contrast and placed little demand on the ability of the monitor for a faithful reproduction. As intermediate luminance values important to the recognition of the stimuli or the response of the participant are added, the risk of falling below the critical 3:1 ratio increases (such that the contrast variation across monitors will begin to confound the study). In designs requiring more complex stimulus in a Web study, variations in contrast should be verified to ensure that the dependent measures in the lab are not altered and to ensure that the variations do not confound the effects of the independent variable.

One method of using intermediate values of luminance with some success on a monitor will be to use them to antialias or smooth an image. Such high contrast steps of going immediately from black to white are beyond the sampling capability of the display in most cases (Krantz, 2000, Silverstein, Krantz, Gomer, Yeh, & Monty, 1990). This is why such lines have these unwanted jagged steps, the jaggies, in them. Making the line wider than one dot and using the extra width to make the edges dimmer can eliminate these jagged steps as well as reduce the problem of pixel independence. This does not reduce the overall contrast of the line but it does reduce the contrast where the jaggies are and renders them less visible to invisible. Ideally, one should gamma correct luminance and use a linear luminance fit to a gaussian curve to make this antialiasing work (Silverstein, et al., 1990). However, for many cases, it is enough to make the jaggies less visible and irritating and any approximation will work fine. The top half of a monitor's gamma is very nearly linear ($r = .99$) so those top steps can be used with fair certainty (Krantz, 1988). This use of intermediate luminance values will have the added benefit of reducing the effect of pixel interaction discussed above. Since the lines are wider than a pixel, they will not have the same degree of pixel interaction (see Figure 3 above).

In summary, to minimize the effect of the monitor do not use real fine stimuli. We get that lesson both from the data to minimize pixel interactions and the need to minimize the jaggies. In addition, it is advisable to use simple stimuli that do stress the ability of the monitor to faithfully reproduce the image.

Instructions to the Participant

I have mentioned above in some places that instructions to the participant may be a way to gain some control over the environment and thus over the experiment's stimuli. In this section I am going to suggest some instructions that can be provided to study participants. The types of instructions discussed below refer to lighting, sizing the stimuli and seating distance from the monitor. It should be noted that there is no guarantee that instruction will be followed. The researcher might ask participants to give some information about lighting and distance as feedback in order to assess whether the participant has paid attention to and followed the instructions.

Lighting

If the contrast is an issue in the experiment, one should suggest that the participant draws the shades and/or turns off overhead lights. The instructions might suggest that if they cannot at least draw the shades they should not participate in the experiment. Light from windows is one of the worst sources of light that can fall on a screen surface and reduce contrast. It is possible for the window to be behind the participant and the light from the window to fall on the display washing much of the image out and rendering the image practically invisible. This source of variation in image contrast needs to be eliminated from almost any experiment, but particularly experiments that require timed responses (Krantz et al., 1992).

Screen Resolution and Browser Size

It is important for the stimulus to be as similar in size across the participants as possible in most cases. However, different users will set their screens to wildly different settings and most graphic formats do not adjust to screen size. If the screen resolution is much lower than expected, say 600 x 480 when 1280 x 1024 is expected, some of the stimulus might be truncated off of the screen. If it is the opposite of the above, the stimulus might be too small to be seen. While one may feel this problem has been resolved by having images scaled (as easily done in Web pages), there are two problems here. First, this scaling is relative to the browser and not the screen. Secondly, for most bitmaps scaling introduces artifacts into the images that are to be avoided. Of course vector graphics like that found in Flash Graphics have less trouble with scaling but the scaling is still relative to the browser and not the screen.

Instructions should begin by instructing the participant to set the screen resolution to some given size. As all monitors can reduce their resolution but older ones cannot increase resolution, it is best to choose a fairly low resolution such as 640 x 400 or 800 x 600. This low value will rise over time, but slowly, as monitors last and many people hold onto monitors for several years. The next instruction should be to either open a new window to give a size relative to the monitor or resize the existing widow to a given size. The guidelines should inform participants that a new window will be opening. There are users in the world, myself among them, which do not like remote Web pages opening windows or resizing browsers without being informed. It is possible using Java to determine the screen resolution here so one can verify if this instruction has been followed.

Distance

The other size of the equation about the size of the stimulus deals with the distance that the participant sits from the screen. If I sit closer or farther it alters the size of the stimulus and stimulus field. The issue is more than just having the stimulus be in front of the participant where it can be seen. We can only see a small region of space with high acuity (Anstis, 1974, Osterberg, 1935). This inability to read wording outside the fovea can be a significant problem. Let me give an example. In my cognitive psychology class I try to run a lot of the online laboratory studies and use the data from these experiments for our class discussions. One I have used is the partial report experiment at the wonderful site, CogLab (Francis, Neath, & Surprenant, 2000), which is a classic replication of the iconic memory study of Sperling (1960). The results did not come out as expected based on that original study. Well, that statement may be too strong, but neither my students nor me had very good memory for the letters even when the row of letters to recall was signalled immediately after letters were removed. In the class discussion, we came to the common experience that few of us could read the top and bottom lines at all. We could only make accurate responses at the short time interval. Their stimuli fill most of the screen and there are no instructions about how far one should be from the screen. As such most of the top and bottom letters are too far from the fovea to be read. To be most precise the participant must be sitting at a certain distance relative to the width or height of the stimulus. This might be a bit much to ask, as it would involve computation, however, this step should guarantee greater control over stimulus size and its relation to the visual systems ability to resolve details. One way to gain some control over the visual size of the stimulus is to use a real rule of thumb. The thumb's width at arms length is about 1 to 2 degrees in diameter. One could put up a standard sized object and have the subject sit back so that the subject's thumb covers the object.

Conclusions

Ideally, the psychological stimulus should be very carefully specified. With the variations within and between monitors, that means careful and tedious calibrations. Thus, it is often found that calibration is not carefully done even in the lab where such calibration is possible (Brainard, 1998; Krantz, 2000; Metha et al., 1993; Olds et al., 1999). Still more difficult is the situation on the Web where calibration in all but the most general terms is impossible. This situation would seem to imply that we cannot conduct fruitful research over the Web. However, as Krantz and Dalal (2000) have found, and other chapters in this book attest, fruitful research can be done over the Web. The contradiction is resolved by limiting the range of stimuli that are used and the types of questions that are asked.

In general, experiments that require fine precision terms of the stimulus are not possible. Examples of such experiments are those measuring discrimination or detectability or even matches cannot be done. Experiments with color, except when color is used very simply and with only a few very different values, also fall outside the range of possible IV for Web experimentation. We simply cannot sufficiently predict the color at another monitor. Even primaries are not fixed and vary more between monitor technologies than within (Silverstein et al., 1990). An example of a successful use of color as an IV would be as a code in a search value where the presence of color is contrasted with the absence. The stimuli that have been successfully put on the Web are simple graphics and text. They are high contrast stimuli that only need to be about the 3:1 ratio that seems to be a useful guideline for limiting the effects of contrast on behavior. This rules out using many different luminance levels. Line graphics will be the most robust. It is also important to pay attention to the size of the stimuli and either know the monitor resolution or ask the participant to change their resolution to some predetermined value. Also, participants should be instructed to sit at a certain distance from the screen so that the stimuli are of a known size. It would be advisable to determine how influential size is to behavior and chose only those behaviors where size is not a significant factor as well.

In this chapter, I have discussed only a few critical aspects of the stimulus. I have not had space to cover such issues as motion or the effect of the narrow field of view with high acuity that is part of the human visual system. If one wishes to use motion or a small or brief stimulus, pilot test these stimuli to see if they will work in the varied conditions likely on the Web. David Brainard's (1998) excellent bibliography provides further reading on the CRT and related display and is conveniently posted on the Web.

Thus, the stimulus can survive the Web, but only for simple stimuli. There are many important questions that can be asked with these stimuli and many important features can be conveyed. If one goes beyond the static simple images recommended here, stimuli should be pilot tested. However, where subtlety of the image requires calibration, a laboratory setting will be required.

References

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research, 14*, 589-592.

Brainard, D. (1998). Tips raster graphics psychophysics bibliography [WWW document]. URL http://color.psych.ucsb.edu/psychtoolbox/bib.html

Bridgeman, B. (1998). Durations of stimuli displayed on video display terminals: (n-1)/f+persistence. *Psychological Science, 9*, 232-233.

Cook, J. N., Sample, P. A., & Weinreb, R. N. (1993). Solution to spatial inhomogeneity on video monitors. *Color Research and Applications, 18*, 334-340.

Cornsweet, T. N. (1970). *Visual Perception*. New York: Academic Press.

Davidoff, J. (1987). The role of colour in visual displays. *International Reviews of Ergonomics, 1,* 21-42.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*, 951-966.

Fallon, A., & Rozin, P. (1985). Sex difference in perceptions of desirable body shape. *Journal of Abnormal Psychology, 94*, 102-105.

Francis, G., Neath, I., & Surprenant, A. (2000). The cognitive psychology online laboratory. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 267-283). New York: Academic Press.

Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist, 15*, 694-703.

Hu, Q. J. & Klein, S. A. (1994). A two-dimensional lookup table to correct the spatial nonlinearity on CRT displays. *Society for Information Display Digest of Technical Papers, 25*, 19-22.

Keating, C., Mazur, A., & Segall, M. (1977). Facial gestures which influence the perception of status. *Sociometry, 40*, 374-378.

Krantz, J. H. (1988). *Personnel protection implementation program: Vision testing* (Technical Report 88SRC26). Minneapolis, MN: Honeywell Systems Research Center.

Krantz, J. H. (2000). Tell me, what did you see? The stimulus on computers. *Behavior Research Methods, Instruments, & Computers, 32*, 221-229.

Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers, 29*, 264-269.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp.35-60). New York: Academic Press.

Krantz, J. H., Silverstein, L. D., & Yeh, Y-Y. (1992). Visibility of transmissive liquid crystal displays under dynamic lighting conditions. *Human Factors, 34*, 615-632.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*, 37-68.

Laugwitz, B. (2001, this volume). Web-Experiment on colour harmony principles applied to computer user interface design. In U.-D. Reips & M. Bosnjak (Eds.), Dimensions of Internet Science.

Lennie, P. Trevarthen, C., Van Essen, D., & Waessle, H. (1990). Parallel processing of visual information. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurophsiological foundations (pp. 103-128). Orlando: Academic Press.*

Metha, A. B., Vingrys, A. J., & Badcock, D. R. (1993). Calibration of a color monitor for visual psychophysics. *Behavior Research Methods, Instruments, & Computers, 25*, 371-383.

Mollon, J. (1992). Worlds of difference. *Nature, 356*, 378-379.

Mollon, J., Covonius, C., & Zrenner, E. (1998). Special Issue: Proceedings of the International Colour Society. *Vision Research, 38*.

Morgan, M. J. (1991). Hyperacuity. In D. Regan (Ed.), *Vision and visual dysfunction, Vol. 10 (Spatial Vision)*, 87-113.

Myors, B. (1999). Timing accuracy of PC programs running under DOS and Windows. *Behavioral Research Methods, Instruments & Computers, 31*, 322-328.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (1999). Effective color CRT calibration techniques for perception research. *Journal of the Optical Society of American, Part A, 16*, 1501-1505.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica, (Suppl. 6)U.*

Penn, C. & Silverstein, L. (1998). FED design factors for optimum viewability in automotive applications. *Society for Information Display Digest of Technical Papers, 29*.

Reips, U.-D. (1995). *The Web's Experimental Psychology Lab* [WWW document]. URL: http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html.

Senior, C., Phillips, M. L., Barnes, J., & David, A. S. (1999). An investigation in the perception of dominance from schematic faces: A study using the World-Wide Web. *Behavior Research Methods, Instruments, & Computers, 31*, 341-346.

Silverstein, L. D., Krantz, J. H., Gomer, F. E., Yeh, Y., & Monty, R. W. (1990). The effects of spatial sampling and luminance quantization on the image quality of color matrix displays. *Journal of the Optical Society of America, Part A, 7*, 1955-1968.

Silverstein, L. D. & Merrifield, R. M. (1985). *The development and evaluation of color systems for airborne applications* (Technical Report #DOT/FAA/PM-85-19). DOT/FAA.

Sivak, M., Olson, P. L., & Pastalan, L. A. (1981). Effects of driver's age on nighttime legibility of highway signs. *Human Factors, 23*, 59-84.

Snyder, H.L. (1988). Image quality. In: Helander, M. (ed.), *Handbook of human-computer-interaction (437-474)*. Amsterdam: North Holland.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs, 74* (11, Whole No. 498).

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Welch, N. & Krantz, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, & Computers, 28*, 192-196.

Wertheimer, M. (1912). Über das Sehen von Scheinbewegungen und Scheinkorporen. *Zeitschrift für Psychologie, 61*, 463-465.

Wiggins, J., Wiggins, N., & Conger, J. (1968). Correlates of heterosexual preference. *Journal of Personality & Social Psychology, 10*, 82-90.

Zeki, S. (1993). *A Vision of the Brain.* Cambridge, MA: Blackwell Scientific Publications.