Significant visual impact: Is it or isn't it?

Arthur E Stamps III, Ph.D., AIA
Institute of Environmental Quality
San Francisco, California 94110
url aestsamps@ix.netcom.com

ABSTRACT

Thirty-eight countries, from all continents except Antarctica, have formal environmental impact review procedures. These impact procedures typically require distinctions between "significant impacts" and "non-significant" impacts. For some issues, such as visual quality, distinguishing the major from the trivial impacts is especially difficult. This paper outlines a theory of visual impacts, shows how the theory can be implemented, and illustrates the theory with three cases histories and a survey of research on the effects of various planning policies. The case histories are examples of statutory and discretionary design review in California and include specifying bay windows on houses, specifying contextual fit, and a before and after study of decisions of a review board. The talk concludes with a discussion of the ranges over which the theory will or will not be applicable and of the opportunities for future cooperative international research.

1 INTRODUCTION

1.1 Needs

Environmental impact assessment is widely required at international, national, and local levels. Key documents such as the Stockholm conference of 1972 and the United States' National Environmental Protection Act of 1969 both call for careful evaluation of the effects of large projects on the environment. Article 130 of the Treaty of Rome calls for the preservation, protection, and improvement of the quality of the environment, and it also calls for the European Community to consider the potential benefits and costs of actions or inactions which might relate to the environment. Today, thirty-eight countries, from all continents except Antarctica, have formal environmental impact review procedures (Glipin, 1995).

One common aspect of environmental impact analysis is the visual effects of proposed projects. In the United States, the National Environmental Protection Act requires consideration of visual impacts. The state of California, a leader in setting environmental policy (Fischel, 1995) has the stronger requirement that public agencies must make findings regarding impacts based on substantial evidence, facts, and a reasoning process (Johnston and McCartney, 1991). California explicitly requires a determination of aesthetic effects in environmental impact analyses (California Council of Civil Engineers & Land Surveyors, 1990, p. 125), and also requires that "The determination of whether a project may have a significant effect on the environment calls for careful judgment on the part of the public agency involved, based on the
extent possible on scientific and factual data." (California Council of Civil Engineers & Land Surveyors, 1990, p. 59).

When applied to individual buildings, visual impact analysis is also conducted under the term "design review". Design review is nearly universal in the United States. In thirty of the states, the courts have established that aesthetic considerations alone are a sufficient purpose for regulation (Smardon and Karp, 1993). Over ninety per cent of the larger cities in the United States use design review (Lightner, 1993). Recommendations for improvement from city officials included making the guidelines more explicit, expanding the jurisdiction of the reviews, consolidating the process, and involving more technically competent staff. In a separate survey of architects' opinions of design review, the main recommendation for improvement was to have technically competent reviewers, leading Lightner to suggest that "...the controversy [regarding design review] often revolves around the question of whose tastes will prevail, who has the right to determine 'good design'" (Lightner, 1993, p. 2).

Decisions about environmental aesthetics typically require distinctions between "significant impacts" and "non-significant" impacts. Projects with "significant impacts" are treated much more stringently than projects the impacts of which are "not significant" (Canter and Canty, 1993), so there is a very strong practical interest in distinguishing significant from non-significant impacts. In the impact literature, the term "significant" means an impact large enough to require consideration and possible mitigation, while "non-significant" means too small to be worth consideration. Distinguishing between these categories of impacts requires at least two judgments: first, an estimate of how large the impact will be, and second, if that impact is strong enough to require attention. Thus categorizing impacts as significant or non-significant assumes that we can measure the strength of the impact.

For environmental aesthetics, the question of how to measure the strength of visual impacts is open and contentious. A decade ago Cats-Baril and Gibson (1986) reported findings from a Delphi survey of a wide variety of people involved in environmental aesthetics. There were 28 respondents from twelve academic fields ranging from agriculture to philosophy, and another 21 respondents from public and private policy organizations, Federal administrators, architects, landscape architects, and lawyers. In their collective opinion, the most important task in environmental aesthetics was the "development of a dynamic, systematic, reliable, valid, & implementable methodology" for measuring visual impacts. Other tasks were inclusion of social and cultural values, legal implementation, economic impacts, and balancing expert and public tastes.

Today (1996), experience indicates that the idea of measuring judgments of environmental aesthetics is still contentious. The practical result is that decisions regarding environmental aesthetics are typically made informally on the personal discretion of decision-makers such as government staff, planning commissions, or
committees of design reviewers or on vague urban design principles. As Lightner (1993) noted, the design review process is typically based on "...banal but occasionally profound rules that are being promoted as good design in cities" and the personal feelings of the reviewers. In other words, design review is highly discretionary, and governmental discretion is a process which is very easy to abuse (Davis, 1971, 1972). However, given the very wide usage of visual impact and design review assessments, one wonders if personal discretion is the best method for making those judgments, or if perhaps other methods would furnish a better design decision and support system for environmental aesthetics?

This paper attempts to provide such an alternate system by listing key issues, outlining a theory of environmental aesthetics, describing implementation protocols, and documenting applications. The first step is to set the agenda by listing the key issues.

1.2 Key Issues

In my own work on environmental aesthetics, I found it necessary to address five key issues. Those issues are as follows:

1. Under what conditions can debates based on taste be resolved? This issue is important because many discussions regarding aesthetics simply bog down between personal declarations of what is or is not beautiful. Decision-makers need systems for making decisions, not for prolonging debate. So it is useful to know the conditions under which decisions about aesthetic issues can be made.

2. Measurement: Can intensity of feelings be determined reliably? This issue is important because many people hold that feelings cannot be measured in the first place, and so measurement is inapplicable to decisions regarding environmental aesthetics.

3. Consensus: Does it matter who judges? Here the issue is whether or not there is sufficient consensus among either people or demographic groups to support a decision based on the public interest. If there is a high degree of consensus, then there will be a well-formed public interest. Converse, if there is not a high degree of consensus, then there will not be a well-formed public interest and the decision-maker will not be able to satisfy everyone with a single design.

4. Visualization: Can we know the visual aspects of a project before building it? This issue arises simply because it is better to evaluate visual impacts before, rather than after, construction. This issue is contentious because it is widely believed that environments cannot be simulated but rather must be experienced directly.

5. Administration: What institutional processes will produce accurate, precise, fair, and efficient estimates of visual impacts? This question arises because the dominant
form of visual impact analysis is done through discretion by government officials, and discretion is prone to being arbitrary and capricious.

We now turn to the development of a theory of environmental aesthetics which may help us find answers to these questions.

2. THEORY

2.1 Individual judgments

As Kant (1791) noted, aesthetics involves judgments of taste based on feelings of pleasure or displeasure. Thus aesthetic judgments necessarily involve feelings. Feelings, Kant goes on to say, are the purely subjective element in a representation. The objective part is the object given in perception. Feelings do not describe what is being judged. Lyrical language and common usage often blur this distinction, as when someone says "This environment is beautiful". If feelings refer to the subjective aspect of the judgment, then the correct way to express the beauty of an environment is to say "This environment makes me feel pleased, enchanted, happy, etc.". A useful heuristic for distinguishing between the subjective and objective aspects of an aesthetic judgment is to try to insert the adjective into the following format: "This _____ makes me feel ___.". If the sentence makes sense, the adjective belongs to a feeling; otherwise it belongs to the object. For example, the phrase "This environment is pleasing." would translate into the sentence "This environment makes me feel pleased.". On the other hand, the sentence "This environment is long." would translate into the sentence "This environment makes me feel long.", which does not make sense. Thus pleasure is a feeling which describes the subject, while length describes the object. So a part of aesthetic judgments is subjective and a part is objective.

The subjective/objective distinction carries over into a distinction between qualitative and quantitative research methods. People who are comfortable with feelings tend to prefer qualitative methods such as narratives about personal lives, histories of art or art movements, or discussions about what makes great art or design excellence, and tend to be uncomfortable with the idea that quantitative methods might be applicable to aesthetics. Quantitative researchers, on the other hand, are usually uncomfortable with the apparent lack of reliability and precision in the qualitative methods, which appear to generate an unstable body of knowledge. I will not pretend to solve this dichotomy here; better researchers than I have tried and failed. (C.P. Snow's Two Cultures are alive and well.) Rather I will try to make the case that quantitative methods can faithfully reflect an aspect of the subjective side of the aesthetic judgment which is useful for determining visual impacts.

The useful aspect is the intensity of the feeling. While there may be any number of objections to any particular way of measuring feelings, I have never met anyone who
denies that feelings can be strong or weak. For pleasure, feeling ecstatic is more intense than feeling pleased, and feeling pleased is more intense than feeling so-so, and I doubt that anyone would disagree with this simple ranking of these three feelings. The simple ranking of these feelings is not usually a matter of contention. However, when a specific quantitative method for measuring the intensities of these feelings is proposed, qualitative proponents are apt to take exception to the method. The method I usually employ is an eight point semantic differential scale of ranging from extremely pleasant to extremely unpleasant. I use semantic differential ratings for several reasons. First, research has shown that many forms of scaling, including comparative choice, rank-ordering environments, placing pictures on a table, and using several psychophysical response functions all produce results which correlate about .98 (Stamps, 1996a). Second, although the semantic differential scaling generates non-parametric data, it turns out that the power of parametric and non-parametric statistical tests is virtually identical down to sub-sample sizes of 10 (Zimmerman and Zumbo, 1993), so use of normal parametric statistics is appropriate. Third, analyzing raw semantic differential data is far easier than any other method, because it just involves calculating average preference ratings (Schroeder, 1984).

It might be noted that this scale of pleasant/unpleasant does not indicate all the information which might go into an aesthetic judgment. For example, one person could rate an environment as being extremely pleasant because of its physical qualities, while another could rate the same environment as extremely pleasant because a member of her family slept there. To be really specific, I loved Jerusalem because of the play of morning light on the pink limestone with which the city is veneered; my mother loved Jerusalem because I was there. These personal reasons would not show up in a semantic differential scaling of preferences for Jerusalem, but, then again, they are not supposed to. For issues of public aesthetics, the useful information is the intensity of the feeling, not the cause of the feeling. In addition, emphasis is placed on the visual aspects of the environment, as contrasted to the olfactory, acoustical, tactile, gustatory, or kinesthetic aspects. The reason for this is that many studies, covering 152 environments and 2400 respondents, have found high correlations (r6, k .04) between preferences obtained on site and preference obtained through visual simulations, so it would appear that, in general, these preferences are based mostly on visual properties (Stamps, 1990). Incidentally, other work, based on a more extensive sample of 1215 environments and 4200 respondents, suggested that preferences obtained for even simple simulation methods such as photomontages or computer models correlate well with preferences obtained from slides (Stamps, 1993a), so it is now feasible to use visual simulations to represent un-built projects.

Another implication of this theory is that the semantic differential ratings do not describe the environment, but only the feelings of the subjects. To continue Kant's distinction between subject and object, parts of the environment should not be specified in terms of subjective feelings, but rather in terms of properties of the physical environment (Stamps, 1994a). The are many ways to describe physical
environments. The method I use is based on Aristotle’s four causes: material, formal, efficient, and final. The material causes are the stuff of which a part of the environment is made: wood, stone, glass, etc. I would classify light as part of the material cause, on the analogy that photons are as much a part of the visual environment as pollen is a part of the olfactory environment, and pollen is definitely a property of plant materials. The formal aspects are specified in terms of Euclidean geometry in an affine space. Each point can be uniquely located. More complex physical arrangements such as distance, scale, outlines, or patterns can be specified as relationships among the specified points. Phenomenological space (how an environment appears from a given place) is specified in terms of projective geometry from a given location. Relationships such as above, below, in front of, behind, enclosed, diaphanous, hidden and even just plain visible can all be specified in terms of projective geometry. Other concepts of projective geometry which are quite useful in visual impact analysis are the station point, angle of vision, occlusion, viewshe, distance, areal extent, and aerial perspective. The efficient cause for a given environmental scene is the process through which a particular environment was created. The major distinction here is typically between natural and built environments, but social, political, or economic conditions would also be part of the efficient cause of an environment. Finally, the final cause is the purpose which the environment serves. Functions such as residence, commercial, industrial, or public are typical final causes.

In this theory, therefore, some parts of aesthetic judgments are attributable to the subjective feelings, and some parts are attributable to objective properties of the environments. The feelings are specified in terms of ratings of pleasure or displeasure, and the environments are specified in terms of their material, formal, efficient or final causes. Being a designer by nature, my own inclination is to learn about how qualities I could create might influence people’s feelings, so the empirical work described in this paper tends to focus on the formal and material causes of environments. It is left up to others to report on the influences of efficient or final causes.

2.2 Collective judgments

So far we have only considered personal judgments of environmental aesthetics. However, for visual impact analysis, judgments of many people are typically involved. When more than one person or one group of people is involved, there is the possibility that different people or different groups of people may have different tastes. For example, the adage that “beauty is in the eye of the beholder” implies that individuals have different tastes. Distinctions between expert and populist taste imply that this demographic distinction is an important aspect of preferences. However, a review (Stamps and Miller, 1993) of the relevant literature suggested another alternative: that there was a very high correlation (r = .86) between expert and popular taste for environmental scenes. That finding was somewhat unexpected, so I ran a series of studies which were designed to find out how much environmental preference could be
attributed to the subjects and how much to the environments. For 3 studies, covering 238 scenes and 139 respondents, it turned out that the environments accounted for about 40% of the preference variance while the respondents accounted for only about 10% (Stamps, 1995a, 1995b, 1996b). For these data, demographic effects were quite small, so the dominant determinant of preference for these scenes was the environment per se.

Another interesting finding from these studies was how much preference variance could be attributed to different orderings of the stimuli. This finding was not obvious since it depended on several technicalities. First, the raw data is composed of ratings of environments by either people or groups of people. A typical data matrix would be ratings (1..8) by nsubj for nstim. Second, the correlation matrix of this data is calculated. Third, the eigen values of that correlation matrix are calculated. Fourth, the eigen values are interpreted as the amounts of variance attributable to independent dimensions. These inferences are all in keeping with the usual interpretations of factor analytic studies. However, there is an additional inference which is not always noted: that each independent dimension in a linear space can represent (and usually does represent) a different ordering of the preferences for the stimuli.

Let us define a preference "viewpoint" as an ordering of items in terms of preference. The implication is that the eigen values indicate how much of the total preference variance is attributable to how many viewpoints. If one eigen value accounts for 90% of the preference variance, then all the subject groups pretty well agree on what is good and what is bad. Conversely, if the eigen values are on the order of 1/n(groups or 1/nsubj, then each group or person will have a different preference ordering. Thus the eigen values of the preference correlation matrix will indicate the degree to which multiple preference viewpoints are needed to explain how much of the preference.

So far this is all theory. What makes it interesting is that, for the three studies for which I had demographic breakdowns, one single viewpoint accounted for over 90% of the preference variance. This implication is that there was very high consensus among demographic groups for the environments studied. I have found demographic differences (especially political affiliation) in other studies of highly contentious projects (Stamps, 1991a, 1991b), but when random sampling is used for both respondents and stimuli, the stimuli appear to be the most important preference determinant.

Getting back to the dialectical aspect of environmental aesthetics, the default rule of thumb is that if aesthetic judgments are completely subjective, then it is pointless to try to resolve aesthetic disputes. Kant (1791, p. 205) discusses this point in his section on the dialectics of aesthetic judgment. Most of us have probably had the experience of having different aesthetic judgments than other people about the same object. In my experience, aesthetic disputes phrased in terms of what is "ultimately" or "really" beautiful are apt to be a waste of time, as are aesthetic disputes based on
correspondences between the object under contention and "great" works of art. In either case, each party to the dispute may select whichever definition of "ultimate", "real", or "great" fits his or her personal taste, and there is no basis for resolving the differences. Another non-converging dialectic can be created by relying on the credentials of the disputants. In design review, for example, the typical case is for the architect to claim to know what is beautiful because of talent, training, and license, while planners tend to base their claims on an unbiased intent to act in the public interest and on the police power.

On the other hand, if we are willing to accept a restricted definition of aesthetic merit in terms of ratings of pleasantness, then resolution of aesthetic differences is feasible. If there is what Kant called a "sensus communis", or consensus of what is aesthetic (Kant, 1791, p. 151), then that consensus will be reflected in the empirical fact that most preference variance is attributable to a single linear ordering. The mathematical version of this idea is that, for a correlation matrix on preference ratings by people by artists, there will be one large eigen value and a bunch of small eigen values. That is what I have been finding in my studies of environmental preferences. For the environments I have studied, the presumed distinctions among personal and demographic preferences has been moot: there has been a single dominant preference ordering — Kant's "sensus communis" — among the environmental preferences. For that data, it would clearly be a waste of time to dispute differences in taste: there weren't any of which to speak. So, if there is a high degree of interpersonal or demographic consensus, then the questions of who judges or who should judge issues of environmental aesthetics become moot.

3. IMPLEMENTATION

3.1 Administrative Implementation

Visual impact and design review analysis is usually done by governmental agencies, and the actual decisions are made by administrative governmental employees. Because these decision-makers are members of the government, their actions are more limited than are the actions of private parities. For example, the US courts have held that the beauty of the environment is a legitimate public interest (Williams, 1974; Duerksen, 1986; Poole, 1987), but that interest should be based on the preferences of the general public, not on the personal tastes of the public officials (Ziegler, 1986). This limitation would seem to suggest that public officials have a compelling need to ascertain public responses to environmental projects whenever a decision regarding environmental aesthetics is made.

Two court cases suggest other needs for public officials who make decisions regarding visual impacts. The first is Nollan v. California Coastal Commission (483 U.S. 825 (1987)). In Nollan, the court held that there must be a connection between a
legitimate governmental interest and what was being taken. In Nollan, for example, the governmental unit wanted the property owners to give a strip of their beach front property to the public so people could walk along the beach. However, the governmental interest that was supposed to be served was to not allow new buildings to block the view of the ocean from a coastal highway. Since the view of the ocean from the highway was precisely the same regardless of whether the strip of beach was public or private, the taking was unrelated to the stated governmental interest. The holding was that there had to be a connection between the stated governmental purpose and the property that was taken. The second case was Dolan v. City of Tigard (94 Daily Journal D. A. R. 8803, U.S. Supreme Court 93-518, decided June 24, 1994). This case carried the line of reasoning two steps further. First, in addition to the requirement of a connection between the taking and the governmental interest, there now had to be a determination that the public benefit obtained was roughly proportional to the private harm caused. Second, it was up to the government to make that determination. In nonlegal words, under Nollan and Dolan, U.S. governments that wish to take private property for public use need to establish the degree of causal relations between the purported governmental interest and the property taken.

The Nollan/Dolan requirements are not yet legally binding on issues of environmental aesthetics in California, but the extension would seem to make a great deal of sense. After all, if the purpose of a public official’s decision is to improve the appearance of the environment, then clearly that public official will need to know what affects the appearance of the environment and how strong those effects are. The extension of the Nollan/Dolan requirements to judgments of environmental aesthetics can be made by phrasing the requirements in terms of the visual impact theory outlined above. The purpose for a determination of an issue of environmental aesthetics is whether a particular project would, if built, diminish, maintain, or enhance the visual amenity of the geographical area affected by the project. The project itself would be defined in terms of specific materials located at set of points in three-dimensional space, and the affected geographical area would be defined as all the points from which the project is visible. The administrative procedures would require creating images of the project as visible from public places, obtaining images of the surrounding environment, and conducting preference experiments to find what changes the proposed project would make in the overall preferences for the environment. If, for example, a project were not visible at all from public places, then the Nollan requirement could not be met by the governmental unit, and so aesthetic regulation would be impermissible. To take another hypothetical example, if the project would produce only a very minor negative change in the overall visual amenity of the area, but the costs of preventing that minor effect were high, then the project would fail the Dolan requirement.

It might be noted the same considerations apply to governmental actions. If a governmental unit suggests modifications to a project, then the government needs to find out (a) if there is any causal connection between its recommended modifications
and the visual amenity of an area, and (b) that the change in the visual amenity is large enough to justify the costs. If, for example, a planning department wanted to chop off a floor of a house to make the neighborhood look better, then it would be up to the government to run the experiments demonstrating that there was a large enough increase in visual amenity to justify reducing a house three to two stories. For a governmental process such as a design review board, the corresponding analysis would be to run experiments demonstrating that the projects were better after review than before review. For urban design guidelines, the appropriate information under Nollan/Dolan would be to demonstrate that building projects which conformed to the guidelines produced better looking environments.

In short, I am proposing that governmental units think of visual impacts as interventions and to use scientific methods to find out how well those interventions work. An intervention is just an action taken to change the existing state of affairs or to redirect a future state of affairs. In gardening, for example, fertilizing a flowering plant is an intervention intended to make the plant have more flowers. In social welfare, a typical intervention would be to conduct an immunization clinic to prevent children from getting preventable diseases. In visual impact analysis, an intervention could be a project, alternate versions of a project, code sections imposing restrictions on physical design components, or a discretionary design review process. For visual impact analysis, the scientific method consists of interpreting those interventions as hypotheses, finding or creating environmental scenes which exemplify or do not exemplify those interventions, conduct preference studies, and find out what effects implementation of the interventions would have on the visual amenity of environments. Use of a random sample of existing buildings would justify claims that a particular intervention would diminish, maintain, or enhance the visual quality of the area, and use of a random sample of respondents would avoid any possibility that a public official was using personal caprice instead of representing the public interest.

3.2 Statistical implementation

In terms of the actual mechanics, the proposed statistical implementation is to use random sampling to obtain subjects and stimuli, to obtain preference ratings from each subject for each stimulus, and use the general linear statistical model to find out (a) how much preference variance can be accommodated by a single ordering of the stimuli, and (b) the standardized mean contrasts among stimuli or groups of stimuli. The amount of preference variance accommodated by a single linear stimulus ordering is the extent to which a sensus communis exists, and the standardized mean contrasts are the estimates of the magnitudes of aesthetic effects. The calculations required by these processes are well documented in the literature (Noble, 1969; Harman, 1967; Overall and Klett, 1972; Winer, Brown, and Michels, 1991).

It should be noted that random sampling on both respondents and stimuli is needed to obtain valid results. Without random sampling on the respondents the data could be
based on self-selected groups with special interests. Without random sampling on the stimuli, each interest group is prone to selecting environmental scenes that most favor their particular interests. Randomization controls for both biases. Randomization is not difficult but it will be tedious. Procedures for respondent and geographic sampling are available in the literature (Deming, 1960, Sudman, 1976, Cochran, 1977).

It should also perhaps be noted that emphasis is placed on standardized mean contrasts, not whether the results achieve a particular level of statistical significance. The reason is that we are interested in the strength of the relationship between preferences and environments. In contemporary statistical practice, that relationship is known as an "effect size" (Rosenthal and Rosnow, 1991). There are many possible ways to measure effect sizes, including standardized mean differences ($d = \frac{\mu_1 - \mu_2}{\hat{\sigma}}$), correlations of any type (product moment, rank order, point-biserial), chi square, proportions, etc. By far the simplest methods are to use the standardized mean difference ($d$) or the correlation ($r$). Both of these measures can be computed on a spreadsheet, so implementation by planning staff is feasible. As it happens, for most applications, both measures can be converted into one another, so researchers or planners can use whichever measure is more convenient for them (Rosenthal and Rosnow, 1991). This paper will focus on the use of the standardized mean contrast ($d$).

Actual implementation of these statistical concepts requires selecting a scaling instrument, images of environmental scenes, and samples of respondents. Since space is limited and since these the technical aspects of these issues have already been discussed in the literature (Stamps, 1994b), I will only summarize the general points in this paper. The general conclusion is that using a simple semantic differential scale of pleasant/unpleasant on slides of environments works well, and that the most efficient method of analysis is a repeated-measures analysis of variance in which $\sqrt{mse}$ is used as the estimated population standard deviation ($\hat{\sigma}$). Contrasts between any desired combinations of stimuli are computed by averaging the preferences from all respondents for each stimulus, and then subtracting the average stimuli preferences from each other. A typical analysis consists of the raw data matrix (nsubj by nstim), an anova table, and a table of contrasts. Examples can be found in Winer et al., (1991). Effect sizes are obtained by dividing the mean contrasts by $\sqrt{mse}$ (Rosenthal and Rosnow, 1991).

One interesting interpretation of effect sizes is that they can be used to distinguish the mountains from the molehills, the material from the trivial, the wheat from the chaff, and the significant from the non-significant visual impact. For simplicity, consider the following case. A preference experiment for a random sample of houses in a neighborhood generated a standardized preference score of 3.0. Two alternatives for a new house were proposed, one of which raised the preference score to 4.0, and one
of which lowered the preference score to 2.9. In terms of effect sizes, the preference effect of the first alternative was +1.0 and the preference effect of the second alternative was -0.1. Which effect would be significant? Here we have to be somewhat careful, because the statistical interpretation of "significant" differs from the environmental impact interpretation. In the environmental impact usage, "non-significant" means too small to be worth consideration. The legal term "de minimus" is perhaps a more appropriate, though non-standard, phrase.

But what value of a standardized contrast would correspond to a "de minimus" effect? Cohen (1988) suggests, as rough guidelines, that effect sizes below .20 are small and are unlikely to be noticed by casual observation. Medium effect sizes (d = .50), would be distinctly noticeable, and effect sizes over .80 would be grossly noticeable. In order to calibrate effect sizes against visual images, I computed standardized mean contrasts for 3221 pairs of slides in my files and ordered them in terms of effect sizes (Stamps, 1996a). Cohen's recommendations appeared to fit the contrasts I had obtained: if the standardized mean contrast were less than .20, it appear to be very hard to decide which slide was the better. On the other hand, if the contrast were .5 or higher, it was quite clear which scene was better. If a specific numerical value is needed to distinguish significant from non-significant visual impacts, I would currently recommend a standardized mean contrast d = .2. Other options are, of course, available. For example, a survey could be run in which residents were asked which buildings in their neighborhood were objectionable. Then a preference experiment could be conducted which contrasted the objectionable buildings to a random sample of buildings in the neighborhood. That experiment would establish a numerical value for distinguishing how bad a building had to be before it became objectionable. Still another option would be to defer judgment on the threshold until it was demonstrated that (a) there was an effect, and (b) it was in the right direction. For example, if a project had no discernible effect, or if the effect were negative, then it would be pointless to continue a discussion about how small an effect had to be before it was non-significant.

In short, I am proposing that questions regarding the magnitudes of visual impacts be resolved through empirical science rather than rhetorical discourse. Standard scientific methods of random sampling, experimental design, and the general linear statistical model will produce accurate and precise estimates of how large and in what direction the visual impacts of proposed projects would be. In order to illuminate how this process works, I would now like to describe some case histories.

4 CASE HISTORIES

The case histories are examples of statutory and discretionary design review in California and include specifying bay windows on houses, specifying contextual fit, a
before and after study of decisions of a review board, and a summary of design interventions for which I have preference data.

4.1 Bay Windows Statute

From the preceding discussion about the legalities of land use planning in the United States, it is clear that aesthetic controls on architecture can serve a legitimate government interest only if those controls result in more aesthetic environments. In this case I investigated the efficacy of one such control. On October 6, 1978, the San Francisco planning code was changed to allow bay windows on the fronts and rears of buildings. Several code sections were involved, of which the most important was Section 136. The general idea was to replace the flat street facades with facades which featured bay windows. Figure 1 shows line drawings of these two styles. Following passage of section 136, a large number of residences which conformed to the code were built. Because the majority of these houses were built in the Richmond district of San Francisco, the houses became known as "Richmond Specials". As more and more of these residences were built, it became more and more obvious that the aesthetics of the resulting houses were quite poor. In order to find out if the presumed disamenity of this code section was real I conducted some experiments (Stamps, 1993b, Stamps and Miller, 1993). The respondents were chosen at random from the population of San Francisco. Randomization was used to ensure that the findings would not be biased in favor of any particular political faction. Then a random sample of houses in San Francisco was photographed. This sample was the control group. Finally, a sample of Richmond Specials was photographed. This was the treatment group. The experimental design was a simple contrast between the control and the treatment group, and the hypothesis was that preferences for the treatment group (e.g., Section 136) would be higher than preferences for the control group (the random sample). In policy terms, the hypothesis was that houses built to conform to the code section would raise the average visual amenity of the city. However, when the data were analyzed, it turned out that the standardized mean contrast was -32, which means that the code lowered the visual amenity of the city. Clearly this code section did not serve a legitimate public interest. Instead of achieving a public good, it was promoting a public bad. The obvious implication are that this code section was counter-productive and the planning department should have validated the code section before implementing it, or, without such prior validation, should been prevented from implementing it.

Insert Figure 1 about here

Figure 1. The pre-code houses were being built before the planning code was changed in 1978. The planning code minimum "Richmond Specials" were the result of the planning code. Sources: Stamps, 1993b, Stamps and Miller, 1993

270
4.2 Contextual Fit Statute

Without the rapid, accurate feedback produced by scientific experiments, it took eight years before the mistake of Section 136 was addressed. On November 4, 1986, there was a general election, and in the general election there was Proposition M. Proposition M was placed on the ballot by citizen's groups as a direct confrontation to the official planning policies. The main purpose of Proposition M was to restrict the construction of high rise buildings beyond the limits proposed by the planning department, but the language of Proposition M had two other important implications. First, by overturning the plans proposed by the planning department, Proposition M shifted planning authority from the planning department to the voters. Second, Proposition M contained language requiring "that existing housing and neighborhood character be conserved and protected in order to preserve the cultural and economic diversity of our neighborhoods." The effect of this proposition was to legitimize opposition to residential development by neighbors or neighborhood groups on the grounds that the construction was incompatible with the character of the existing neighborhood.

Subsequent regulations extended Proposition M by requiring that construction in residential neighborhoods be compatible in both scale and character with surrounding areas. Experimentally, the design heuristic of "matching the scale and character" held up much better than the design heuristic of "add a bay window". In one experiment (Stamps, 1993c) I made photomontages of blocks of houses. There were ten blocks, and each block had houses of a single style (High Victorian, Mission, stucco box, etc.). The highest rated blocks had the High Victorians, and the lowest had stucco boxes. Then I photographed seven three story High Victorians, seven two story high Victorians, seven three story Stucco Boxes, and seven two story Stucco Boxes. These twenty-eight photographs were pasted up into blocks in which the houses were either mixed or matched in scale or character (Stamps, 1994c). Overall, people liked blocks in which the houses matched the scale and character of the adjacent houses more than they liked blocks in which a house contrasted in either scale or character (d = .46). Other results were that people liked the homogeneous blocks the most, and it did not matter if the houses were all small and plain or all large and fancy. The worst-liked blocks were the blocks in which only one building was different. Again, it did not matter if the different building was larger or smaller, only that it was different. Inserting a High Victorian into a block of Stucco Boxes lowered the preference for the block. Figure 2 illustrates some of the main findings. It appeared that the design heuristic of "matching the scale and character" had some actual validity in terms of increasing the visual amenity of these residential blocks.

---

Insert Figure 2 about here

271
Figure 2. In a study of scale and character, people had definite preferences which depended on how well a new building fit with its adjacent buildings. Source: Stamps, 1994c.

The validity was not, however, complete. Many of the permit applications for residential construction in San Francisco involved the addition of a third story to a two story house. Under the "scale and character" theory, both scale and character were important design components. Yet, in a study which controlled for character and which explicitly tested standardized mean differences for two and three story houses, it turned out that there was virtually no difference in preference between these two scales ($d = .03$), (Stamps and Miller, 1993).

4.3 Citizen's Discretion

Concurrently with the passage of Proposition M, some of the more active neighborhoods began to establish their own local design review boards. One such board was established in the Bernal Heights neighborhood in 1986. Membership on the nine member board was restricted so that no more than one full member and one alternate could have any connection to or training in architecture. None of the members was paid, so this was a design review committee of amateur lay personnel. The process was quite complicated. On the average, it took over a year for a house to move through the neighborhood review process.

This case is of more than local interest because it was a direct manifestation of the policy of delegating power to the people. People whose philosophies follow Rousseau and De Tocqueville would probably be inclined to support the neighborhood review board, while people who followed Hobbes would probably be inclined to question the review board. In either case, the delegated power was the authority to perform design review under the police power of the state, and the question was whether this delegation had beneficial effects on the visual amenity of the neighborhood. In scientific terms, the question was whether the operations of the review board had beneficial effects on the proposed designs.

In order to address this question I conducted several empirical preference experiments. The first experiment was done on line drawings of houses as originally designed and as after the houses had been modified to address the recommendations of the review board. The respondents were the members of the review board and 135 residents of San Francisco (Stamps, 1991a). The second experiment was conducted after the houses were built, and used photographs of the as-built houses (Stamps, 1992). Examples of the houses are shown in Figure 3. The results were quite clear. Overall, the review board had a small beneficial effect ($d = .17$). The correlation between the review board's preferences and the preferences of San Franciscans for the as-built houses was also beneficial but small ($r = .10$, $d = .20$). Conversely, the correlation
between public preferences for the design drawings and the as-built houses was .77 (d = 2.13). These findings have two important implications: (a) the preferences of this review board, even though it was composed of lay members, did not reflect the preferences of its presumed constituency, and (b) simple experiments, based on line elevations, was a much, much more accurate method for determining preferences for houses. In terms of policy implications, it would appear that the time and effort expended in this neighborhood design review process were unlikely to be justifiable under the Nollan/Dolan criteria.

Figure 3. Preferences by a lay neighborhood review board were weakly related to other people’s preferences, but pre-construction experimental preferences by other people were strongly related to post-construction experimental preferences. Source: Stamps, 1991a.

4.4 Survey of planning interventions

In addition to studies described above, I have conducted studies on some other design interventions. Table 1 lists those interventions and their effects on the visual amenity of their environments. Again, the general idea is to think of each planning principle as a design intervention, run a preference experiment illustrating scenes that do or do not exemplify that intervention, and calculate the effect size of the intervention. The first intervention compared houses which matched the scale and character of adjacent houses against houses which contrasted either scale or character (but not both) of adjacent houses (Stamps, 1994c). The effect size was .46. Next in efficacy was the design principle of adding trees to a row of houses. In this experiment (Stamps, in submission), blocks of houses were created with or without trees, and the standardized mean contrast between the blocks with trees and the same blocks without trees was .35. The third principle was the efficacy of lay design review. This study compared houses as originally submitted for lay design review to houses as modified to respond to the lay design review requests. The standardized mean contrast fell to .17, indicating a positive but small effect (Stamps, 1991a). This finding was replicated in a study comparing houses which passed professional design review to houses which were exempt from professional design review (Stamps & Nasar, in submission). In this study the standardized mean contrast was .14, indicating another positive but small effect. The next principle was that deleting wires from a residential block would improve the visual amenity of the block. The experimental finding supported this principle but at relatively low efficacy (d = .12). Principle 6 was that adding a third story to a two story house would decrease the visual amenity of a block. The efficacy of that principle was very low (d = .03) (Stamps and Miller, 1993). Eliminating cars from street scenes had an even smaller effect (d = .01, Stamps, in submission). Finally,
### Table 1. Effect sizes for some planning principles.

<table>
<thead>
<tr>
<th>PRINCIPLE</th>
<th>MSE</th>
<th>CATEGORIES</th>
<th>NSTIM</th>
<th>( \mu )</th>
<th>EFFECT SIZE (( \delta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Contextual urban design principles</td>
<td>1.19</td>
<td>Match scale and character of adjacent houses</td>
<td>4</td>
<td>5.51</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contrast in either scale or character of adjacent houses</td>
<td>8</td>
<td>5.01</td>
<td></td>
</tr>
<tr>
<td>2. Trees</td>
<td>0.91</td>
<td>With</td>
<td>12</td>
<td>5.04</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Without</td>
<td>12</td>
<td>4.70</td>
<td></td>
</tr>
<tr>
<td>3. Lay design review</td>
<td>1.29</td>
<td>After review</td>
<td>6</td>
<td>.66</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Before review</td>
<td>6</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>4. Professional design review</td>
<td>2.29</td>
<td>Passed</td>
<td>7</td>
<td>5.29</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Was Exempt</td>
<td>7</td>
<td>5.07</td>
<td></td>
</tr>
<tr>
<td>5. Wires</td>
<td>0.91</td>
<td>Without</td>
<td>12</td>
<td>4.93</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With</td>
<td>12</td>
<td>4.81</td>
<td></td>
</tr>
<tr>
<td>6. Height of houses</td>
<td>1.58</td>
<td>2 Story</td>
<td>16</td>
<td>4.26</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Story</td>
<td>1</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>7. Cars</td>
<td>0.91</td>
<td>Without</td>
<td>12</td>
<td>4.88</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With</td>
<td>12</td>
<td>4.87</td>
<td></td>
</tr>
<tr>
<td>8. Residential urban design code of 1978</td>
<td>1.58</td>
<td>Random Sample of City</td>
<td>8</td>
<td>4.12</td>
<td>-.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Code Complying</td>
<td>8</td>
<td>4.53</td>
<td></td>
</tr>
</tbody>
</table>

Sources for table 1 are given in the text.

Houses exhibiting the urban design principles exemplified in the San Francisco Planning Code of 1978 (Section 136) were compared to a random sample of existing houses in the city (Stamps & Miller, 1993a). The resulting contrast was negative and substantial \((\delta = - .32)\), indicating that this particular set of urban design principles was counterproductive. These data suggest that local design interventions of contextual fit and planting street trees had large, positive effects, that design review and eliminating wires had small effects, that regulating a third story or eliminating cars had a trivial effect, and that implementing the 1978 bay windows code had a negative effect. Extensions to other planning principles, policies, or procedures should be straightforward.

### 5 DISCUSSION

#### 5.1 The proposed model

The proposed basic model of visual impacts presupposes that judgments of aesthetics have both subjective and objective parts; that the subjective part is the intensity of
feeling a person has in an environment, that the objective part is composed of attributes of the environments, that the intensities of the feelings can be expressed in terms of simple semantic differential ratings, that the attributes of the environments can be specified in terms of Aristotle's four causes and in terms of Euclidean and projective geometry, and that estimates of the strengths of the relationships between feelings and environmental attributes can be determined through the general linear statistical model in general and through standardized mean contrasts in particular. The standardized mean contrasts provide numerical indicators which distinguish among deleterious visual impacts, trivial visual impacts, and beneficial visual impacts.

We can now provide some preliminary answers to the five key issues described in Section 1. First, debates regarding taste can be resolved if there is a way of expressing the intensities of people's feelings, the qualities of the objects under consideration, and the populations of people and objects to be evaluated. For judgments by individuals this process will probably diverge, but for issues of public aesthetics, where there are many people involved, the available data indicate that there will be rapid convergence.

Second, simple semantic differential ratings appear to indicate the intensities of people's feelings quite well, so estimation of overall aesthetic effects is feasible.

Third, judging from the available data, there appears to be substantial consensus regarding the aesthetic merits of environmental scenes. Thus the public interest in environmental aesthetics has been, so far, well-formed and would provide a solid basis for public policy decisions. Consequently, for the scenes reported in the literature, and for respondents chosen at random using scientific protocols, the question of who judges does not seem to matter very much. It is only until the respondent population is restricted to a small group or to self-selected parities, such as members of review boards or special interest groups, that demographic distinctions appear.

Fourth, accurate and inexpensive visual simulation techniques are now available and provide the material with which the visual aspects of unbuilt projects can be expressed.

Fifth, the preliminary evidence suggests that discretionary review processes are unreliable and may actually be counter-productive. The scientific protocols proposed in this paper are suggested as an alternative procedure for evaluating the aesthetic impacts of proposed projects or governmental policies and procedures.

The data also suggest that separating "significant" from "non significant" effects may not be the most appropriate way to analyze visual impacts. For one thing, it may not be necessary. If a project is invisible from a public place, the effect will be zero, and so the government has no legitimate interest in regulating the appearance of the project. On the legislative side, if a governmental regulation decreases the visual amenity of an area, then again no public interest is being served. Only if the effect is in the right direction will the question arise of just how small is too small.

275
Mathematically, the problem of distinguishing "significant" from "non significant" impacts is that the effect sizes of visual impacts are clearly continuous, while the categories of "significant" and "non significant" are discontinuous. The transition from continuous to discontinuous variables can be accomplished in various ways, including setting a threshold based on another criterion or by minimizing a combination of the visual impact effect size with another criterion such as cost. For example, if data were available indicating that houses would not be objectionable unless they were at least .8 σ below the average preference for the area, then a threshold of $d = .8$ would distinguish significant from non significant visual impacts. In my judgment, after looking at over 3000 pairs of slides, a $d = .8$ implies a very large visual effect, that d's outside the range of -.2 < d < .2 would provide better estimates of "significant" visual impact.

An alternate procedure would be to obtain measures of visual impacts for various alternatives, calculate the costs associated with each alternative, and select the option with the highest benefit/cost ratio. For example, in the cases of residential construction, the public benefit would be measured in terms of the standardized mean contrast between two or more alternatives and the random-sample control group of the area. Costs would be measured by construction cost, resale value, or life-cycle income estimates. Then benefit/cost ratio's could be calculated and the alternatives with the highest ratio identified.

It is still an open question whether these findings generalize to other interventions or to other cities. However, after twenty-years' experience, I have found it to be most unwise to challenge either Gauss's theory of errors or the central limit theorem. In my opinion, if there is a sensus communis on an issue of environmental aesthetics, the experimental protocols proposed in this paper will find it very quickly and very efficiently. Technical staff whose mission it is to support policy decisions about visual impacts or environmental aesthetics will find these protocols to be accurate, precise, and efficient methods for discovering the information they need to do their job.

5.2 Implications

One simple implication of the present findings is that governmental units should have the burden of finding out how large an effect their requirements would have on proposed projects before the requirements are implemented. This is not a particularly new idea: the Federal Administrative Procedure Act of 1946 (Gellhorn, 1981, pp. 331–407) requires that any new rule by a Federal Agency be validated before being implemented. What is new is the idea that validation of aesthetic judgments is possible and feasible and therefore is now necessary for good public administration regarding environmental aesthetics.
Another implication of these findings is that disputes regarding environmental aesthetics can now advance beyond rhetorical discourse and personal caprice. Neither rhetoric nor personal caprice support converging dialectics, and so an infinite amount of time and effort can be expended on effects which are non-existent or trivial. On the other hand, the scientific protocols support a converging discourse, in which the trivial can be separated from the material, and attention can be focused on the things that really matter. As governmental units face more and more stringent budgets, and as there is more and more pressure to increase the quality of their work, the methods outlined in this paper will be more and more suitable to their missions.

A final implication of the current findings is that future international cooperative work on environmental aesthetics would be most efficient if everyone reported their work in terms of effect sizes. A large number of researchers currently collect ratings data, but report their findings either in terms of non-metric scatter plots or in terms of whether or not a particular finding achieved a significance level below .05. Use of non-metric plots prevents the application of probability theory. However, given the results of Zimmerman and Zumbo (1993), the non-metric simplification is unnecessary, and so parametric statistics can be used instead. Use of the .05 significance level confounds the effect size with the sample size, and so prevents synthesis among different studies (Rosenthal and Rosnow, 1991). However, if we all use common metrics such as standardized mean differences or correlations, then data from Europe can be synthesized with data from Australia or from the United States, and we can all save ourselves a considerable amount of time and effort by mathematically incorporating each other's findings into our own conclusions (Hedges and Olkin, 1985). After all, free work by someone else is the most efficient of all possible experimental designs, and so reporting results in terms of $d$ values is a very good way to support the development of a reliable, rapidly-converging collective body of knowledge on environmental aesthetics.

6 SUMMARY AND CONCLUSIONS

It is suggested that issues of environmental aesthetics be addressed with the scientific protocols of random sampling and statistical inference. Photomontages are valid and are currently the most cost effective simulation medium. The magnitude of visual impacts can be measured as standardized mean contrasts ($d$) between groups of stimuli. $d$'s less than 0.0 indicate a deleterious project or governmental requirement. Current data suggest that $d$'s in the range of ±20 are likely to indicate a trivial project or governmental requirement, while $d$'s outside the range of ±5 or more will indicate a substantial effect.

In terms of the legal requirements implied by the Nollan/Dolan decisions of the U.S. Supreme Court and the data reported in this article, projects with a visual impact of $d$ greater than -2 with respect to a random sample of the existing area should be
considered to have non-significant impacts and so should be exempt from visual impact or design review. If the contrast is less than -2, a project would be deleterious, and governmentally-suggested modifications might be justifiable. On the other hand, governmental requirements with d's less than +2 would have trivial or deleterious effects, and hence would not be legal. Only modifications to projects with d's less than -2 or governmental regulations with d's over +2 would serve a legitimate government interest in environmental aesthetics.

6 REFERENCES


Pre-code Houses

Planning code minimum (Richmond Specials)
People liked

Victorians when seen by themselves

Blocks with houses that matched either the scale or character of the other houses

Blocks with several large buildings

more than they liked

Victorians in a row of boxes.

Blocks with houses that contrasted in both scale and character with the other houses.

blocks with only one large building.
Review board preferences, based on pre-construction drawings

Other people's preferences, based on pre-construction drawings.

Other people's preferences, based on as-built photographs.