Testing Prospect-Refuge Theory: A Comparative Methodological Review

Annemarie S. Dosen¹ and Michael J. Ostwald¹

¹The University of Newcastle, Callaghan, Australia

ABSTRACT: In 1975, Jay Appleton theorised that certain spatio-environmental conditions engender feelings of safety and security by providing opportunities for outlook and seclusion. This proposition, which has since become known as prospect-refuge theory, has been widely debated in a range of fields and has found an enduring place in environmental psychology. It first became popular in architecture in 1991 when Grant Hildebrand demonstrated its relevance to Frank Lloyd Wright’s housing design. Hildebrand used this theory to argue that particular combinations of spatial qualities – including enclosure, light and view – are essential for a liveable space. While prospect-refuge theory has since been widely used to interpret a range of architects’ works, and to support many theories of design, only limited empirical evidence is available to substantiate the theory. This paper analyses thirty separate attempts, over a twenty-eight year period, to examine the validity of prospect-refuge theory. The paper is not concerned with the results of these tests, but rather with the methods that have been used, their characteristics, strengths and weaknesses. The purpose of this paper is to provide a comprehensive, critical foundation for future empirical or sociological studies of this important theory as it relates to architecture.

Conference theme: Architecture and environment
Keywords: Prospect-Refuge Theory, Habitat Preferences, Design Assessment

INTRODUCTION

In 1975, British academic Jay Appleton published The Experience of Landscape (1975) in which he describes the human behavioural and psychological need to inhabit places that allow a person to observe without being seen. Appleton traces this desire for safety to Charles Darwin’s late nineteenth century anthropological beliefs about human survival instincts. In addition, he also postulates that under certain conditions, pleasure can be derived from the observation of particular environments. To support this assertion, he recalls Dewey’s (1934) proposition that beauty is a result of the interaction between an individual and an environment rather than being an innate property of some object, place or view. Furthermore, Appleton proposes that it is the combination of certain characteristics of the view and the observer’s enclosure that generates this aesthetic and emotional response. For example, if a view is able to evoke or suggest sublime feelings (‘peril’ and ‘hazard’) then this can contribute to a positive experience of space. Fundamentally, Appleton’s (1975, 1984) research maintains that the right combination of outlook or prospect and enclosure or refuge will produce an environment that has a clear and beneficial impact on individual spatial preference.

Since being proposed, prospect-refuge theory has been widely discussed by art historians, philosophers and spatial psychologists and in the 1990s Grant Hildebrand applied it to the residential architecture of Frank Lloyd Wright. Hildebrand (1991, 1999) expanded the theory adding further qualities including formal complexity and the opportunity for spatial exploration. In the years that followed, architectural scholars, designers and critics have repeatedly made reference to prospect-refuge theory and some architects – including Wright, Aalto, Murcutt, Kundig and Zumthor – have become strongly associated with this theory. However, despite its recurring presence in architectural theory and design, there is only limited empirical evidence of its efficacy and most of that is only indirectly related to architecture. In order for this theory to have any validity for architects further research should be undertaken which has a specific focus on aspects of design. Nevertheless, before this can occur, past approaches to examining prospect-refuge theory must be revisited as a precursor to considering how future research should be undertaken.

There are various ways of developing evidence for or against a proposition. For research about human reactions and responses, the most common methods are derived from the social sciences. Such methods are innately idiosyncratic; they are concerned with the development of evidence for a position from the rigorous analysis of human values, beliefs and behaviours. Quantitative studies undertaken in the social sciences are typically founded on a philosophical position which in turn supports the application of an investigative strategy and a framework for evaluation of any data collected (Groat and Wang 2001, Naoum 2007). In combination, these three factors make up the research “method”. In practice, social sciences approaches rely on surveys, interviews and observational studies. Such studies gain their validity through the belief that it is possible to extract a statistically reliable indicator of the reaction, responses or opinions of a larger group of people from a much smaller sub-set of this group. However, in order for this to be true, the research method, including the design of any survey instrument, must be especially carefully conceived.
In the context of research into prospect-refuge theory, Arthur Stamps III has reviewed a range of qualitative and quantitative studies concluding that about half of all research into this topic “used methods of humanistic speculation or personal experiences” (2006, 643) whereas only about 11% were based on empirical data. In the former category, approximately 50% of the research is comprised by the works of Hildebrand (1991), Heerwagen and Orians (1993) and Roberts (2003). In the undefined 39% of works are a range of mathematical studies of isovists and computational models of spatial experience (Franz et al. 2004, Stamps III 2006, 2008a, 2008b). It is the last group, the 11% of empirically based studies using social sciences methods, that the present paper is focussed on. Specifically, this paper undertakes an analysis of thirty studies of the validity of prospect-refuge which were conducted between 1983 and 2010. The present paper is not concerned with the results of these studies, but rather with the methods that have been used. Thus, the research compares the strengths and weaknesses of survey and observation techniques, which rely on a range of media and on different evaluation strategies. In this paper these studies are categorised and compared, before they are in turn assessed for their viability for supporting architectural claims about prospect-refuge theory. This analysis, the first of its type in architecture, provides a critical overview of how this theory has been previously tested while providing advice on how it might be examined in the future.

This paper is structured in two major sections, the first (1.0), which is largely descriptive, is divided by the focus of the individual studies and the second (2.0), which is more analytical, is divided by methodological characteristics. In part one the focus of the research is categorised into four types of environments: (1.1) landscape, (1.2) natural and built environment, (1.3) urban environment and (1.4) interiors. In part two there are five subsections which describe different approaches to methodological issues: (2.1) the questionnaire, (2.2) type of stimuli, (2.3) viewing conditions, (2.4) participants and (2.5) stimuli content. Both parts one and two consider the same thirty studies, but categorised and analysed in different ways.

1. METHODOLOGIES

1.1. Prospect-refuge theory in landscape

In the earliest of the studies recorded in this paper, Nasar et al. (1983) argue that we experience our exterior environment intermittently over time. Thus, for investigating prospect-refuge theory, Nasar’s team interviewed pedestrians who had, over a long period of time, experienced selected sites on a university campus. These sites were classified as (a) open view with a protected or unprotected observation point and (b) closed view with a protected or unprotected observation point. A balanced number of 30 male and 30 female students were selected (15 on each of the four sites) and each assessed the view from one observation point, rating it on a nine bi-polar adjective scale questionnaire which was prepared by eight judges (graduate students of City and Regional Planning). Some of the bi-polar assessment criteria used include, safe/unsafe, interesting/non-interesting and attractive/unattractive.

Five years later, Ruddell and Hammitt’s (1987) study of prospect-refuge related phenomena examined visual preference for edge environments: including the boundaries of a lake, a river, a forest and a meadow. The purpose of their study was an exploration of functional orientation in the first field and not, explicitly at least, “to test the validity of [prospect-refuge theory] for explaining edge preference” (1987, 252). In their research, 160 visitors to an American state park were asked to rate 32 black and white photos along a 5-point Likert scale ranging from “liked not at all” (1) to “liked very much” (5). The 32 photos were selected from a set of 150 images taken of the state park (no details were provided about how) and were presented in a questionnaire as prints (5 x 7.5 cm). In their paper, the researchers proposed five edge environment themes which correlate to prospect-refuge concepts (including “prospect dominant”, “refuge dominant” and “refuge symbolic”) in relation to the viewer’s position and the degree of the distinct edge. A “refuge dominant” scene, for example, was composed of two edges, one in the foreground and the other in the background and a “prospect-dominant” scene contained open-field scenes with a very distant treeline. Seven of the 32 photos showed edges where the forest-meadow interface was unclear (categorized as “rough ill-defined” environments).

Figure 1: (1) Refuge-dominant, (2) Prospect-refuge balanced, (3) Prospect-dominant.

Rachel Kaplan’s and Eugene J. Herbert’s (1988) study of landscape preference focuses on the influence of familiarity with landscapes in a cross-cultural context. Their research was conducted in both North America and in Australia using rural landscape photos taken in Michigan (USA). According to the researchers, both cultures where their surveys were undertaken – Michigan and Western Australia - have similar life styles and economic issues but the vegetation in Western Australia is distinctly different from that of rural Michigan. In their photographs, no people or close-up views of houses were shown. 97 students from The University of Michigan and 122 students from The University of Western Australia (all psychology students) were asked to rate 55 slides using a 5-point Likert preference scale. Three example slides were used at the beginning to demonstrate the process and the rating form.

46th Annual Conference of the Architectural Science Association (ANZAScA), 2012, Griffith University
was numbered to 60 to avoid an “end-effect”. The participants were gathered in small groups, two random orders were used to present the slides and each slide was shown for 10 seconds. There is no further information about the type of the slides (colour or black and white) or the gender and age of participants.

Janet Conrad’s (1993) dissertation examining prospect-refuge theory in design describes four different tests: a pilot study and experiments 1, 2 and 3. The pilot study is focussed on a particular methodological issue: whether artificial images are useful for exploring environmental preferences. For this study Conrad prepared a pencil drawing of Mt Egmont with a lake next to a forest along with some foreground. She then created seven variations of this image adding elements, typically called “symbols” by Appleton, associated with enhanced prospect and refuge characteristics (such as a view framed by a tree in the foreground; a fence across the middle area; an open gazebo in the foreground). Two groups of 23 participants (trained or untrained in design) took part in the study and rated paired images for preference on a 7-point Likert scale. After the pilot, Conrad’s (1993) first formal experiment sought to examine if designers, who are not aware of prospect-refuge theory, can intuitively construct pleasant or unpleasant scenes (using a model stage set with a defined range of “props” representing prospect and refuge symbols) as well as those who are aware of the implications of prospect-refuge theory. These scenes were then rated by two groups of design students (trained or untrained in prospect-refuge theory) on a 6-point Likert scale from (1) “least pleasant” to (6) “most pleasant”. A second part of this first experiment expanded the method to investigate if participants with five years design training and a knowledge of prospect-refuge theory would be able to construct prospect-refuge dominant and prospect-refuge balanced settings and if participants with basic design skills but no knowledge of the theory would identify balanced images as pleasant and dominant scenes as unpleasant.

In her second experiment Conrad (1993) used eight Australian scenic photographs and eight coloured line drawings (traced from these photographs) to investigate if responses to both media would correlate and if preference would be linked to prospect-refuge qualities. 50 participants were recruited who formed three groups and were divided into two subgroups to rate either the photos or the drawings. 40 of the participants were members of a fitness studio varying in age from 18 to 60. Half of this group rated the stimuli in order of preference while the others were asked to rate them on a 6-point Likert scale from unpleasant (1) to pleasant (6). Then, the ten participants of the trained group rated the stimuli for prospect and refuge symbolism on a 7-point Likert scale. For her third and final test Conrad (1993) asked 12 children aged seven (year two of a primary school, seven boys and five girls, from culturally diverse backgrounds) to create pleasant pictures on a flannel graph (with green fabric for ground and blue for the sky) using “props” such as trees, hills, water, sun, boats, shelters and animals. With this test she sought to investigate the innateness of prospect-refuge theory which, according to Appleton, is derived from “human nature which is communicated innately from generation to generation” (1975, 234).

Caroline M. Hagerhall (2000) analysed traditional Swedish landscapes seeking a connection between aesthetic preference and factors proposed both in prospect-refuge theory and in the Kaplan’s (1982) framework for prediction of preference. Hagerhall’s study used eight criteria which include preference, exploration, safety, overview, physical age along with others associated with cultural heritage and evolutionary aesthetics. 119 participants (58 male, 61 female, aged between 21 and 80 years) were selected from a phonebook (excluding Non-Swedish names, 1200 subjects were contacted). The participants were asked to rate 60 colour slides, which were chosen from a sample of 600 photos (depicting sunny or clear days and which excluded figures, water and man-made objects), on a 5-point Likert scale ranging from “not at all” (1) to “very much” (5). The slides were rated for only one question at a time to avoid being influenced by the rating of another criterion and two example slides introduced each question. The questions and slides were presented in two different orders and each slide was shown for 10 seconds.

1.2. Prospect-refuge theory in urban design

Nasar (1988) argues that the aesthetic quality of an environment influences the sense of being attracted to, or repelled by, certain spatial conditions which in turn have an impact on human well-being. Nasar’s study is a comparative analysis of visual preference for urban street scenes in Western (United States) and Asian (Japan) cultures. Nasar assumed that novel scenes would be preferred over familiar scenes and that preference was expected to be relative to visual complexity. His criteria for evaluating prospect and refuge include novelty, complexity, order and openness. His study is based on a daytime-view of urban environments as seen by a driver on a main road of four American and four Japanese cities which differ in density, age and structure. Also, the participants’ education differs and only a few Japanese students operated a car, most using subways and trains. 29 Japanese graduate students (only one female, 18 from architectural engineering, 11 from environmental engineering) and 17 American graduate students (five female, from City and Regional planning backgrounds) participated in the study. All of the participants were between 20 and 30 years old. The students were asked to rate eight, three minute-

Source: Conrad 1993, unpublished Master-Thesis

Figure 2: (1) Strong refuge, (2) Prospect-refuge balanced, (3) Prospect (original image).
videos and 24 slides. The videos were usually recorded on the middle lane at a speed of 30 km/h. The colour slides were taken every ten seconds. Three slides were selected at random and similar slides were replaced by another random slide. Every scene was pre-assessed by expert judges (the researcher and three students with training in architecture but from different cultural background) against a 7-point bipolar adjective scale (including closed / open, chaotic / orderly). The participants also used a 7-point bipolar scale to evaluate each video and slide (pleasant / unpleasant / interesting / uninteresting).

Loewen, Steel and Suedfeld (1993) used prospect-refuge theory as a foundation to analyse the relationship between urban environments and the degree to which people feel safe from crime. In their study, they asked 55 undergraduate students from a psychology class (36% male, 64% female, mean age 21) to generate a list of environmental features that are relevant to such a safety. In their second study, 16 slides were presented to another group of 100 students (45% male, 55% female, mean age 20) showing the three most common characteristics identified in the first study (including light, open space and access to help). Before the test, five judges rated each slide for the absence or presence of each feature. In the slides no people or recognizable features were included. The three characteristics were combined in eight different ways showing an urban outdoor scene. Perceived safety was assessed by a questionnaire using a 5-point Likert scale ranging from “not at all safe” (1) to “very safe” (5).

Mumcu, Duzenli and Ozbilgen (2010) analysed spatial features of successful seating areas in urban open spaces. 148 pedestrians in the Hungarian Turkish Friendship Park (Trabzon, Turkey), were interviewed about their preferences. The participants (63 male, 85 female, aged between 16 and 55 years and from different backgrounds) were asked to highlight on a plan where they would prefer to sit, or not to sit, if they had the freedom to choose. In addition, they were also asked to rate on an eight item questionnaire the places they had chosen using a 5-point Likert scale ranging from “absolutely I do not agree with” (1) to “absolutely I agree with” (5). The qualities the researchers sought to study include prospect and refuge symbols along with consideration of the “most attractive”, “comfortable” and “safe” areas that “enable overview” and a “variety in activity”. Prospect was subdivided into “wide, unimpeded view” and “easily see what is going on around”.

1.3. Prospect-refuge theory in natural and built scenes

Stamps III’s (2008a, 2008b) research into environmental preference is based on a series of large scale studies of the following four factors: “prospect (depth of view), refuge (presence of protective regions in front of the observer or […] possibilities of escape), direction of light (either front or back lighting), and venue (natural or built environment)” (2008a 147). Stamps III uses two broad scales, “comfort” and “liking”, to evaluate spatial responses. Comfort represents perceived safety or danger and liking refers to those results chosen for purely aesthetic reasons.

In his first test (2008a) Stamps III created static colour computer renderings of scenes both with and without refuge. The presence of refuge was suggested by the position of a permeable object (bush, tree, railing) in the foreground. Also, daylight conditions were considered by simulating light coming from both behind the observer and in front. 29 participants (24 male, 5 female, mean age 22, all students but with no further information about background) were asked to rate 16 landscape scenes on an 8-point Likert scale ranging from “uncomfortable” (1) to “comfortable” (8). Two example slides were used at the beginning and every scene was shown for 20 seconds. A second related experiment considered the impact of prospect, which was defined as depth of view (close, medium, far), while refuge was indicated by occluding edges (0, 1, 2 or 3) in the scene. 18 participants (balanced in gender, political affiliation, background, mean age 49 years) were selected by a commercial research firm to rate 16 landscape scenes and 14 rooms on an 8-point Likert scale ranging from “dislike” (1) to “like” (8).

In his next paper (2008b), Stamps III reported the results of four new studies on environmental perception and preference. These four studies were based on extensive previous work (277 studies with 41,000 participants) and in them Stamps III focuses on the efficacies of the five factors: prospect, refuge, light, venue and spatial transition. The correlations of the factors with responses of preference or comfort represent the value of “efficacy”. In the first test, rendered images were presented on screen showing three natural and three built scenes but with no human figures. Each image was rendered twice, the second time only with half the brightness to represent the factor, light. The depth of view (three distances each for natural and built scenes) represents prospect. 16 participants (balanced in gender, from different political backgrounds, unfamiliar with the site and with a mean age of 44) who were selected by a commercial research firm, rated the randomised scenes on an 8-point Likert scale from “uncomfortable” (1) to “comfortable” (8). In the second test, the same 12 background scenes and same participants were used but the factor ‘light’ was replaced by ‘refuge’. For this study, the scenes were shown through two different room types which framed the images: one room was a hexagonal shape in plan with higher ceilings, glazed walls and glazed roof (higher and brighter) while the other was a rectangular room with glazed walls (horizontally framed, and darker). The images all showed the same human figures in the same room. The presentation order was different in the second test: the images were also randomized but half of the participants saw the images without refuge stimuli first and vice versa. In Stamps III’s penultimate experiment, 12 scenes of meadows with birch trees were shown at four distances (80m, 60m, 40m and 20m) from the observer to investigate the combined influence of prospect and refuge. Refuge was indicated by having no barrier in the foreground or by having a fence or a view from a window bay. 17 undergraduate engineering students (one female, mean age of 22) rated the images shown as slides in random order on an 8-point Likert scale from “uncomfortable” (1) to “comfortable” (8). As before, two example slides were shown and each slide was presented for 20 seconds. In this test refuge was related to comfort. The final experiment was a replication of the third one, with the same stimuli being used but a different group of 16 participants (balanced by gender, mean age of 44) recruited by a commercial survey firm.
1.4. Prospect-refuge theory in interior design

Suzanne Scott (1993a) describes research undertaken into the relationship between environmental attributes and preference to provide design direction for architects and interior designers. 309 college students from psychology and journalism classes at a Midwestern university (Wisconsin) participated in the study. They rated 80 black-and-white slides of interior spaces by eight categories of visual attributes in terms of preferences on a 5-point Likert scale from “strongly dislike” (1) to “like very much” (5). The slides were shown for eight seconds in random order and nine example slides were presented to the groups of 10 to 49 students. The participants were not familiar with the scenes and they were not educated in arts or design, but there is no information about their age or gender ratio. Scott explains that this is deliberate because past studies of landscape preference have shown that preference is not related to gender or age (Balling and Falk 1982, Lyons 1983). The photo sample was prepared by four expert judges (interior designers) who had selected the scenes from a sample of more than 150 photos by identifying and grouping them by design variables including “spaciousness”, “spatial organisation” and “complexity of visual field”. Also, settings with plants and natural light were evaluated. The photos showed lounges, corridors, meeting rooms, offices and cafeterias of public and commercial buildings, all from similarly-styled contemporary buildings to limit the influence of isolated aesthetic preference.

In a second study Scott (1993b) examined the relevance of “complexity” and “mystery” as predictors of preference for design attributes of interiors. Her study was intended to manipulate complexity and mystery within interior environments and thereby create more desirable spaces. 80 black and white photos were presented to three different groups of students from different levels and degrees to compare preference ratings (group 1, 321 participants: 113 male, 208 female, degrees other than art and design) to the ratings of two independent groups (31 participants each from design classes) who each judged scenes according to assessments of either complexity or mystery. The photos showed public, institutional and commercial interiors such as cafeterias, libraries and lobbies. A 5-point Likert scale was used for each assessment. The ratings of group 1 were based on an eight second exposure to slides. Group 2 and 3 were asked to sort the photographs to achieve a reliable, comparative judgment.

Franz, von der Heyde and Bülthoff (2004) describe an approach to investigate the experience of architectural interiors using surveys and virtual reality simulations correlated with isovists and visibility graphs. Isovists are viewshed polygons that represent the volume of space visible from a given observation point. Quantitative descriptors of isovist polygons, such as area and length of open or closed edges can then be derived from these polygons for comparison with survey results. Graph analysis allows spatial characteristics of an entire environment to be described by comparing multiple positions within a space. Franz et al. (2004) derived 33 characteristics from the isovists to test them as predictors of experiential qualities of interior spaces using the virtual environment of an art gallery. Two groups of eight university students each (aged 20 - 25) were asked to rate 360° panoramic images of interior scenes which were presented in random order. The test consisted of a pilot stage of 18 scenes and of a main stage of 16 refined scenes. The participants had 30 seconds to freely explore the whole room of a scene from a fixed observation point by using a trackball. In the first stage the rating categories were shown as pairs that exhibit particular spatial qualities (including spaciousness, openness, complexity) and more emotional responses to this experience (including pleasure and beauty). In the second stage the two rating categories were separated and shown consecutively; also, the visible room area was more restricted and there were paintings shown on the walls. For the visibility analysis, the set of isovist points generated on a 50 cm grid were measured. For the statistical analysis the results of the eight participants per stage were averaged and compared with the 33 characteristics.

In 2006, Stamps III reported on two further empirical studies of prospect and refuge correlated against perceived comfort for interior spaces. For his first experiment he used computer-generated images of Japanese Tatami mat rooms in two sizes (4.5 m x 4.5 m and 2.7 m x 2.7 m) and with two different ceiling heights (2.7 m and 4.5 m). For the second experiment he used virtual Roman baths in Egyptian Tombs (2.5 m x 1.71 m, 2 m high, and 7 m x 3.4 m, 4.3 m high). In both tests light walls varied (light to dark and vice versa) and transitions from big to small areas (or vice versa) were considered and a view through a window (indirect prospect) of a lake was offered. In the second test the proportion of high to low rooms (or vice versa) changed to wide and narrow rooms (or vice versa). Altogether, 97 participants, who were recruited by a commercial research firm, assessed computer generated images of the rooms. In the first test, 14 participants (balanced in gender, political affiliation and employment, mean age 51) were asked to rate eight slides on an 8-point Likert scale from “uncomfortable” (1) to “comfortable” (8). The results were not statistically significant and so Stamps III increased the sample size for the second test to 83 participants. First two unscaled examples were shown, then, the recruited students (68 male, 15 female, mean age of 20, no further information about their background) were asked to rate eight slides from “uncomfortable” (1) to “comfortable” (8). The slides were shown until all had responded which took about 20 seconds.

2. DISCUSSION

2.1 The questionnaire

In most of the studies, participants were asked to rate stimuli on the basis of either preference or comfort (or some combination of the two). In general, surveys were used to record the respondent’s point of view. Only three of the 30 tests described in this paper used a bi-polar scale and 24 used a Likert scale of which half of this latter group were even scales (6 or 8 point) and the remainder uneven (5 or 7 point). An uneven scale allows participants the capacity to provide a neutral response: this is sometimes useful in a study where there is a desire to reduce emotional stress associated with questions. However, when asking about environmental preference the possibility of a neutral response can reduce the value of the data and should not be offered. In contrast, an even Likert scale effectively forces the participants to make a choice (whether this reflects their true position or not) and delivers at least a
moderate result. Furthermore, the number of options in the Likert scale is significant as some respondents do not like straying from moderate results and are, therefore, much less likely to select extreme options. Thus, a 6-point Likert scale is less likely to elicit results of 1 or 6, potentially limiting the range and usefulness of the survey. Several of the studies note potential problems with Likert scales and surveys of this type including the so-called “end-effect” wherein the latter answers in the survey tend to be more rushed and consequently inaccurate. A further potential problem of these surveys is that by asking many questions in the same way respondents may begin to answer in a pattern which is mirrored by the question. The subject image or stimuli that are reversed (contraponere) can help to avoid such a pattern but it can also cause confusion. Overall, a questionnaire with an even Likert scale and which offers more choices (8-point) is ideal for ratings for preference. The need for contraponere phrasing and the artificial management of the end-effect is more debatable. Randomisation of image stimuli may mitigate the statistical impact of both of these factors.

2.2 Type of stimuli

The decision to use real or virtual stimuli images is nearly balanced across the complete set of studies. In the former category, two studies took place on-site using real-world conditions, whereas 10 used photographs of sites. In the latter, virtual category, 11 studies used computer simulations while four used drawings or photos of crafted scenarios. Three of the studies compared drawings and photos concluding that the use of photos is preferred but there was less agreement about whether these should be in colour or black-and-white. Stamps III has shown that results derived for static colour simulations correlates highly with results obtained on-site. Scott (1993a, 1993b), on the other hand, used black-and-white images to avoid biases related to hue differences and she refers to Acker’s and Kuller’s studies (1972) which have shown that deep contrasts are more consistently related to preference than hue. The main advantage of colour photos is their semblance of reality. Colour attracts attention and it increases memory retention for natural scenes (Wichmann et al. 2002). This insight may influence the choice of stimuli for environmental preference ratings rather than the argument for black-and-white images being considered neutral. Some colours may evoke positive or negative associations and influence ratings, but this seems less significant for environmental studies. When comparing virtual with real stimuli, Franz et al. (2004) refers to previous studies (de Kort et al. 2003) that claim that the key requirement for the use of any simulation is a high pictorial quality such that they may achieve evaluations, similar by the questions for real or the environment’s. They claim that image or environment conditions, flexibility, interactivity and a high perceptual realism are the advantages of virtual environments which are used to simulate architectural space for testing and they refer in their paper to Bülthoff’s and van Veen’s study (2001). Fundamentally, virtual stimuli allow the researcher to create a more controlled study while the use of photos of real-world requires a much bigger data set as well as careful pre-selection. This latter dimension can be problematic because not all studies clarify if the stimuli were preselected by expert judges or by the researcher. 10 tests used stimuli selected by judges while one test (Hagerhall 2000) specifically avoids any form of pre-selection. In summary, for future studies of prospect-refuge characteristics in architecture the ideal stimulus is likely to be: full-colour virtual images that offer a degree of realism but can be prepared under highly controlled conditions.

2.3 Viewing conditions

Several studies have described the physical conditions – room size, viewing distance, light levels, time of the day and presentation orders – under which the stimuli were viewed. However, according to Stamps III (2006), most viewing conditions can be disregarded as the same results can be achieved with different participants, locations and scaling methods. One of the more relevant conditions appears to be the time a person is allowed to view an image before rating it and whether this needs to be controlled or not. Some studies have presented photo stimuli for a specific, fixed time (typically 8 or 10 seconds) while others left participants to use as much time as they needed. The logic behind this is that a presentation time which is too short can stress the respondent and force a poor, fast decision, while a long time frame increases the length of the study and may attract fewer participants. Despite these overarching concerns, the fixed time limit is also clearly a factor of the larger scale testing environment, with big groups of people in lecture theatres watching slides on a wall, rather than being given individual sets of photographs to consider. None of these conditions may be completely relevant for a future study because it is more common today for such surveys to take place entirely within a web-based environment. In such an environment it is possible for an upper time limit to be set (say 15 seconds) but allow for participants to progress more quickly if they desire.

Figure 3: Percentage of tests by (1) type of stimuli and by (2) age of participants.

2.4 The participants

The number of participants is one of the most significant factors for the validity of the survey outcome. The number of participants needed for a valid research project initially depends on the type of survey. Usually, a subset of a group
of people from a certain population of interest is selected to be surveyed. A wisely selected group, coupled with an appropriate survey design, can deliver accurate results with a relatively small number of participants. Usually, 400 is considered a good number of respondents and for surveys which are based on yes/no responses or for which a 50/50 outcome is considered (Data Analysis Australia 2009). In 12 of the 30 tests analysed here, small groups of participants (15 to 29) were asked to rate stimuli and in only three studies were very large groups (200 or more) used. Interestingly, all three studies with very large groups of participants also used large sample sizes (Kaplan and Herbert 1988, Scott 1993a, 1993b). Large groups of participants were typically students while groups of participants which were recruited by a commercial firm were of small size. Some of the studies differentiate participants by gender ratio or age and a few used additional classifiers, like design skills, educational background or political orientation. Only 14 of the 30 studies describe the gender ratio of which six are balanced (50%), one is almost balanced (55-51%), three are slightly dominant (65-56%) and four are dominant (66% or more). Over 60% of the investigated studies surveyed young participants and only one study recruited seniors. A comparison between preference ratings of young participants to the ones of middle-aged and senior respondents is therefore desirable. Furthermore, it is advisable to avoid surveys which are related to familiarity with stimuli (Scott 1993b). Only two of all studies considered cross-cultural aspects (Nasar 1988, Kaplan and Herbert 1988). Surprisingly, none of the studies examined the environmental origin of participants such as the area in which they grew up or where they lived during the past ten years (rural, suburban or urban) which may have a significant impact on environmental preference and on the perception of space. Despite some claims to the contrary, a group which is relatively balanced in gender and age, and with a mixture of backgrounds, is desirable for a future study, however, the key elements for a study of architecture and design may still relate more to education and training.

2.5 Content of stimuli
To classify the content of stimuli is one of the most sensitive and controversial aspects of a survey. Some studies chose a range of largely style-less or bland contemporary designed spaces for the built environment part of their stimuli. Others avoided including any human figures or overtly recognisable features (landmark buildings or landscape elements). Several studies attempted to remove any seasonal or time-of-day affects in their stimuli, while a smaller number used differing lightly conditions in the stimuli as a means of checking the consistency of results. The logic behind such decisions is to limit the impact of either design, lighting or human recognition, and thereby to focus the participants’ attention only on the relationship between outlook and enclosure. The issue that really needs to be considered for this category of the method is what the study is seeking to achieve? If the purpose is to generally test prospect-refuge theory, then attempts to minimise aspects of the stimuli depicted is reasonable. However, if design itself (the material, formal, phenomenal or tectonic qualities of a room or setting) is being tested prospect-refuge theory in relation to design or architectural aesthetics.

CONCLUSION
As this paper demonstrates, prospect-refuge theory has been empirically tested in many studies of which a small number are relevant to architecture, but so far there is only limited evidence for the theory as it is applied in design. Not only is there relatively little evidence, but much of the empirical data that has been collected is potentially methodologically flawed by being reliant on small sample sizes, or groups that are dominated by a single gender, age group or culture. Taking into account the analysis undertaken in this paper, the ideal methodological factors for a future study of prospect-refuge theory in architecture would be:

- a randomised, even-numbered Likert scale questionnaire with more choices (8-point) and more (easy to understand) questions to avoid “end-effect” responses;
- full-colour, virtual images that offer a sense of realism but which can be modified in a highly controlled way;
- online-surveys (rather than printed questionnaires) with a time limit of 15 seconds per image but that allow for participants to progress more quickly;
- a (medium to large) group of participants, balanced in age and gender, who are from diverse cultural, local and educational backgrounds;
- stimuli that depict interior settings suitable for prospect-refuge theory but that do not include recognisable landmarks or human figures.

Furthermore, it is also possible to speculate, based on the review of the 30 previous research projects, that a valuable study of prospect-refuge theory in design might use a consistent plan geometry (room shape) and outlook or view, but carefully vary the following factors:

- interior style: perhaps comparing a 1890s “arts and crafts” style interior with a 1930s modern or a 1980s post-modern interior;
- portal geometry: comparing the impact of different window fenestrations and proportions;
- tectonic conditions: comparing the impact of different colours, textures, materials and construction techniques.
A comparative study could also be constructed using examples cited in past phenomenological research (like Hildebrand’s) to examine whether there is any correlation between the empirical evidence and the traditional humanities-based interpretation of a building or space. Finally, it is notable that only one of the tests of prospect-refuge theory presents a parallel mathematical analysis of the space being analysed alongside the analysis of people’s perceptions of that space (Franz et al. 2004); this too could be a highly productive topic for future research.

REFERENCES


46th Annual Conference of the Architectural Science Association (ANZAScA), 2012, Griffith University