*Psicológica (2012), 33*, 695-710.

# The use of Factorial Forecasting to predict public response

David J. Weiss*

*California State University, Los Angeles*

Policies that call for members of the public to change their behavior fail if people don't change; predictions of whether the requisite changes will take place are needed prior to implementation. I propose to solve the prediction problem with Factorial Forecasting, a version of functional measurement methodology that employs group designs. Aspects of the proposed new policy are factorially manipulated within scenarios, and respondents typical of those whose behavior would need to change are asked to project how they would react. Because it is impractical to validate the projections by seeing if they correspond to what eventually happens, I advocate evaluating validity by invoking a coherence criterion.

For some fifty years, functional measurement methodology has been valuable in elucidating cognitive processes (Anderson, 1996). It must be acknowledged, however, that interest in the methodology is largely confined to a small subset of the academic community. To broaden that interest base, I propose to heed the plea that George Miller (1969) addressed to his colleagues, to "give psychology away". Miller urged psychologists to find domains in which their expertise can be useful and to make their ideas accessible to non-experts. I make the same plea to functional measurement researchers. This paper illustrates how functional measurement might be

* Acknowledgement: This research was supported by the United States Department of Homeland Security through the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) under award number 2010-ST-061-RE0001. However, any opinions, findings, and conclusions or recommendations in this document are those of the author and do not necessarily reflect views of the United States Department of Homeland Security, or the University of Southern California, or CREATE. I wish to thank Richard John and Heather Rosoff for insightful suggestions regarding the manuscript. Correspondence address: David J. Weiss, Department of Psychology, California State University, Los Angeles, 5151 State University Drive, Los Angeles, CA 90032. Email: dweiss@calstatela.edu.

given away by suggesting a practical extension that can be used in the political arena to guide policy decisions.

The task I envision for which functional measurement is suitable is that of predicting public response to a proposed change that calls for people to modify their behavior. The change might be a new law that attempts to regulate an action, or a construction project that encourages people to adopt new ways of handling a recurrent need. An example of the former might be a proposed ban on smoking in particular facilities; an example of the latter might be the construction of a bike path. Before such plans are enacted, the policy maker needs to know whether people will alter their behavior in the desired manner. If behavior is unaffected, then the change is wasteful at best.

History has shown the need for this kind of prediction. The classic example is the 18th amendment to the United States Constitution, ratified in 1919 to prohibit the "manufacture, sale, or transportation of intoxicating liquors within, the importation thereof into, or the exportation thereof from the US." The supporters of Prohibition had a strong moral stance, but they also addressed anticipated tangible benefits to society. These included reductions in crime, domestic violence, and disease as well as increases in worker productivity. However, Prohibition failed; people continued to drink alcohol. The 18th amendment was repealed in 1933 by ratification of the 21st amendment, in part to remedy two unintended negative consequences, namely the rise of organized crime with attendant police corruption, and the loss of federal and state tax revenue from the sale of alcohol. The "Noble Experiment" had proven to be a policy debacle. The expected behavioral changes did not occur, and respect for the law had declined – an undesirable side effect.

The name assigned to the new flavor of functional measurement is "Factorial Forecasting". The label captures the idea that factorial designs are at the core. The hope is that prior to putting their visions into practice, policy makers will routinely invoke the method to predict how their constituents will respond to the proposed changes.

Factorial forecasting employs familiar experimental methods. The stimuli are scenarios in which variants of the proposed new policy are presented. Scenarios with factorially manipulated political content appeared in the early days of functional measurement (e.g., Anderson, Sawyers, & Farkas, 1972). The variations are constituted to reflect a factorial structure that captures the new policy's key components. Respondents are recruited to represent the folks whose behavior will need to change when the policy is in place. Each respondent is exposed to a paragraph containing a particular combination of levels of the factors. I illustrate how the method might have been used prior to implementation of a controversial policy, the construction of the Los Angeles subway.

In the 1970's, Los Angeles traffic was heavily congested, and the sky reeked of pollution most days. A suggested solution was to get the cars off the freeways by converting drivers to subway passengers. Civic pride – most other world class cities have subways – also played a role in the campaign. Obstacles were that the subway would be very expensive to build, in part because earthquake safety standards had to be met, and that acquiring rights-of-way entailed compensation for very some expensive real estate. Still, the authorities plowed ahead, and over the course of several years, the subway was built as the central component of what will eventually constitute an extensive metropolitan rail system.

Unfortunately, the subway has not met its objectives. Ridership is low, and likely consists primarily of people who would be bus passengers if there were no subway (Weikel, 2010). The freeways are still packed during ever-expanding rush hours. Could this disappointment have been foreseen?

Our hindsight recommendation would be to carry out a study whose participants were people who commute daily to central Los Angeles. Respondents would read a scenario describing the proposed subway system, and then answer simple questions about what they would do given the circumstances described to them. The questions would be the same for everyone, but the circumstances would differ according to the cell of the design to which the respondent had been randomly assigned. Two obvious candidates for relevant factors are price and convenience. A typical question might look like this: "Suppose there were a fast, safe subway that had stations located within a half mile of your home and a half mile of your workplace. If each ride costs $3.00 and daily parking near your office costs $4.00, on average how many days per week would you take the subway?" The investigator could set up levels as shown in Table 1.

Additional factors thought to be relevant to the appeal of the subway can be embedded within the scenarios if desired. The analyst might, for example, expand the analysis of convenience by specifying various distances from station to workplace. Additional factors that characterize subjects, such as sex, age, or economic status, can also be added to the design. Negative incentives to discourage driving or subsidies for particular subgroups of subway passengers might also be included as factors.

**Table 1. Layout for 4 (Convenience) x 3 (Price) factorial design**

| | Price per ride | | |
|---|---|---|---|
| **Distance between station and home** | $2 | $3 | $5 |
| .5 mile | | | |
| 5 miles with park and ride | | | |
| 10 miles with park and ride | | | |
| 15 miles with bus or light rail | | | |

Each subject is assigned to one cell of the design and provides one score. In some studies, responses might be nominal rather than numerical; whether the factors affect projected behavior can be assessed in those cases as well (Weiss, 2010). An important advantage conferred by numerical responses is that cell means can be displayed graphically. For policy guidance, it is likely that effects need to be large enough to be readily apparent. Statistical confirmation is of lesser importance.

A possible outcome from the subway forecasting study is shown in Figure 1. The additive pattern, expressed via the parallel lines, suggests that price and convenience both influence projected ridership, and the two factors trade off – low price compensates for inconvenience to some extent.

The possible outcome shown in Figure 2 tells us that price doesn't matter, at least within the limits explored. Convenience is everything.

Convenience is also highlighted in the possible outcome shown in Figure 3, where only a station within a short walk of home could attract the commuter. Once the car is started, the entire trip is by auto. In this case, the convenience of the subway in effect operates as a dichotomous variable rather than a continuous one.

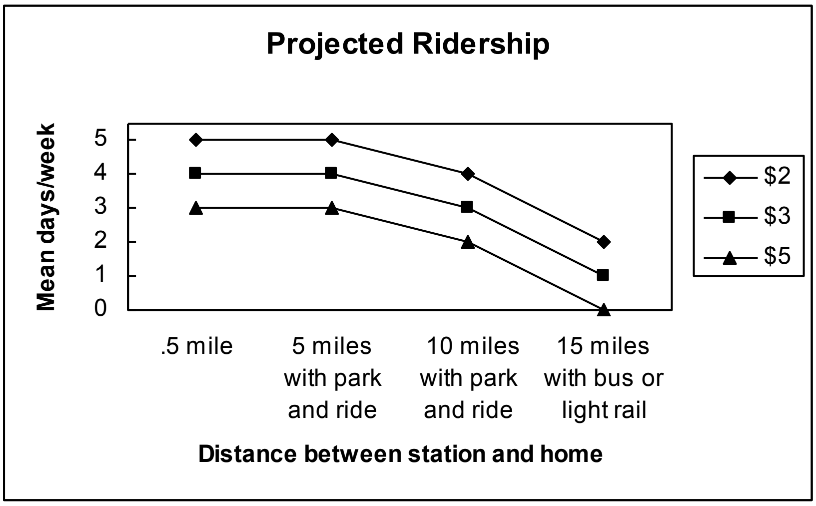


**Figure 1: Hypothetical results from scenario study of projected subway ridership. Each curve displays the mean responses for the indicated price.**

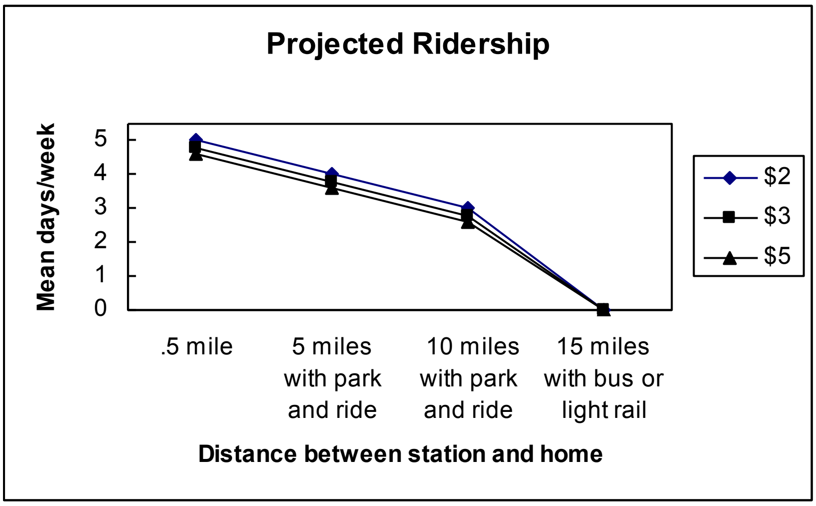


**Figure 2: Hypothetical results from scenario study of projected subway ridership. Each curve displays the mean responses for the indicated price.**

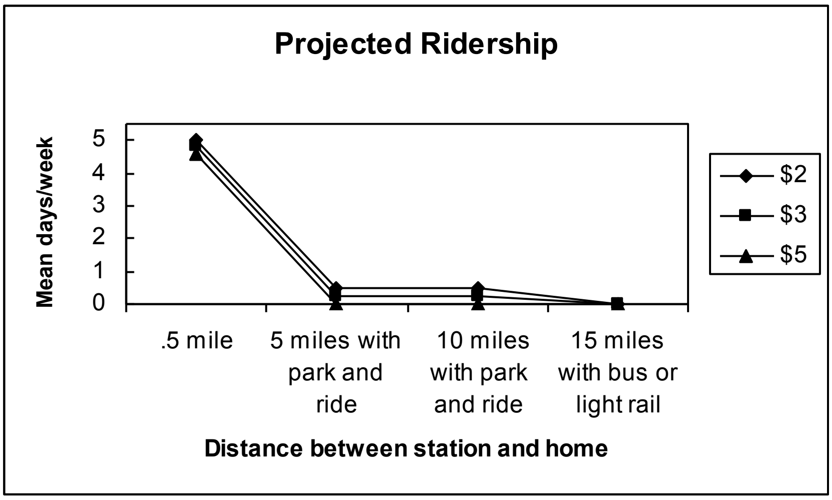


**Figure 3: Hypothetical results from scenario study of projected subway ridership. Each curve displays the mean responses for the indicated price.**

How would these possible results inform the policy decision? If we assume that building a station is a huge expense, then most suburban commuters will not have one near their homes and will have to drive to a park and ride station. Our interpretation of these hypothetical results is that if Figure 1 obtained, building the subway makes sense only if prices can be kept low. Subsidies would likely have to be considered, bolstered by the rationale that reducing the number of cars in the central city makes the entire area more livable.

If Figure 2 were to obtain, the subway would be viable only if many huge park and ride stations could be built 10 miles apart throughout the suburbs. Price does not affect ridership, so the expense of building those stations might be recouped by charging high prices.

In contrast, if Figure 3 obtained, we would suggest that the subway not be built. Commuters would switch to the subway if there were stations everywhere, as there are in London or Paris; but presuming that to be economically infeasible for the huge metropolitan area of Los Angeles, we would be pessimistic about ridership. The practical value of Factorial Forecasting is that it can provide this kind of policy guidance prior to implementation, before the expense is undertaken.

It is important to note that the evidence in the graphs shows much more than attitude statements would reveal. In all three figures, there is at least one condition under which people assert they would be riders. Asking people about attitudes, such as whether they favor building a subway, does not address the details. People might well infer conditions they consider optimal and respond positively. That need not imply they would be riders under the conditions that ultimately come about. In addition, people might express support for a subway because they want to get other drivers off the roads. Los Angeles voters did favor the subway, but hindsight tells us that many of them did not anticipate being riders themselves. If the researcher is able to elicit honest projections of people's personal behavior, it will reveal which policies will be successful and which will not. Of course, a forecast will be more appealing to the policy maker if it does find conditions under which the proposed policy will be effective.

## GROUP DESIGNS

Group designs have been little used in functional measurement research, with the exception of pioneering work by Edmund Howe (1991) and two studies conceived by Christine Rundall (Rundall & Weiss, 1994, 1998). There are several good reasons for preferring the usual single-S designs. It is inefficient to incur the cost of recruitment and then collect only one response from a person. Also, respondents may use a rating scale idiosyncratically, so exposing each person to the full range of stimuli allows individuals to calibrate themselves without confounding scale usage and judgment. From the statistical perspective, single-S designs offer the potential for powerful tests of the hypothesized model, because differences between people do not contribute to the error term.

In the prediction context, however, group designs demonstrate advantages from two perspectives. First, because we are trying to predict group behavior, generalizability is enhanced by recruiting samples from the community whose behavior is supposed to change under the new policy. The inherent power disadvantage of a group design may be overcome by enlisting large samples, though sufficient power is not guaranteed.

Second, we want to avoid contrast and sequence effects. A respondent exposed to the best possible combination of levels will inevitably feel that other combinations are inferior, and might not endorse a combination that would otherwise have been acceptable. This risk may be exacerbated when the stimuli are vivid and memorable, as they would usually be in the policy setting.

With each participant offering only one response, idiosyncratic scale usage is a potential problem inherent in the independent groups design (Birnbaum, 1999). It is therefore best to avoid ratings, although self-anchored scales (Hofmans & Theuns, 2010) may mitigate the problem if ratings are unavoidable. Rather than asking about attitudes or emotions, I suggest employing questions whose answers are cardinal numbers or specific statements about anticipated behaviors. These are less likely to yield responses biased by differences in scale usage.

If the researcher considers carry-over effects to be unlikely, designs in which people provide responses to some but not all stimulus combinations may be feasible. One approach is to use a fractional factorial design (Weiss, 2006) in which respondents are assigned randomly to a subset of the design. Another option is the nested group design (Rundall & Weiss, 1994, 1998), in which a classifying factor determines the subsets. The advantage of these designs is that measurement error can potentially be reduced by extracting consistent differences associated with individuals.

The group design does alter the usual interpretation of the cognitive model. Rather than viewing the model as localized within an individual, the model and the extracted parameters are seen as those of a typical member of the population from which the participants were sampled. The population may be decomposed by incorporating demographic factors such as age, sex, income, etc., within the design, thereby yielding politically tailored predictions.

## RECRUITMENT ISSUES

Who should be the respondents? The stakeholders for a policy decision may include those who build a facility and those who finance it, but the only folks whose responses matter for predictive purposes are the expected users. Statistical theory tells us that the best way to predict is to gather a random sample from the population of interest. However, truly random samples are hard to amass; inevitably some of the designated respondents are unreachable or refuse to participate. In practice, it is likely good enough to recruit people from the relevant group who express willingness to share their perspective. The recruiter must be careful about accessing only those who strongly favor one position or another. If an exercise facility for seniors is under consideration, recruitment should not be limited to people already participating in other exercise programs, but rather should be aimed at as broad a representation of seniors as possible. Biased recruitment leads to erroneous predictions of usage rates.

Once the policy maker has targeted the respondents, how can they be induced to participate in the study? Extrinsic rewards such as cash or movie tickets can be effective incentives, but are not always necessary. The best recruiting pitch may be to truthfully inform folks that their input is important in formulating policy (Weiss & Weiss, 2002). Costs can be kept low by gathering data via private web sites; only those who agree to participate are given access. There are software tools available for mounting scenario studies on web sites; the tools preprocess the data as well. As internet usage becomes increasingly prevalent, reliance upon the web becomes less likely to filter out particular segments of the population.

The skill of the scenario researcher lies in getting the respondents to take the task seriously. It is imperative to make the stimuli engaging, concrete, and specific, so that the respondent envisions being in the situation. Playing a role is more predictive of future action than the fruits of hard thinking about the future (Green & Armstrong, 2011). Questions about intentions (e. g. Brengman, Wauters, Macharis, & Mairesse, 2010), particularly if a moralistic tone seeps through, such as asking whether a person will comply with a medical regimen, donate to a good cause, or complete a task on time, run the risk of inspiring responses biased toward social desirability (Epley & Dunning, 2000). Although the new policy is presented as fictional, it should not take the respondent to an impossibly hypothetical place. For example, a scenario in which the respondent is asked to imagine having far more (or less) wealth or having incurred a life-altering disease is likely to be too fantastic to yield meaningful answers. Respondents who merely go through the motions, working perhaps solely for an extrinsic reward that accompanies participation, are unlikely to provide meaningful data.

## DISCUSSION

In the 1960's, two quite similar methodologies were developed, both based on the idea that examining how variables act in concert could provide a basis for validating judgments. Functional measurement and conjoint measurement (Luce & Tukey, 1964) have always been somewhat contentious competitors (Anderson, 1971; Krantz & Tversky, 1971; Shanteau, Pringle, & Andrews, 2007), although in my view their similarities greatly outweigh the differences. Undeniably, conjoint measurement has been more successful; the name is more widely known and it has received many more citations in the literature. I attribute this popularization primarily to the work of Paul Green (Green & Srinivasan, 1978; Green & Wind, 1971). Green's conjoint analysis is an offshoot of conjoint measurement that proved to be useful in marketing research. Conjoint analysis studies usually ask the respondent to select the preferred product from among a set of factorially manipulated alternatives. Choosing or ranking are held to be easier, and perhaps more natural, tasks for the respondent than rating several alternatives (Louviere, Hensher, & Swait, 2000). This simplicity is seen as a practical advantage for conjoint measurement over functional measurement and other methodologies such as multiattribute scaling (Gardiner & Edwards, 1975) that call for multiple numerical responses.

In asking for projections, Factorial Forecasting avoids comparative judgments. Instead it eases the respondent's burden by asking for only one response in a completely randomized design, or possibly for a few responses in a fractional or nested design. Limiting the number of judgments from an individual also caters to the dark possibility that a respondent may attempt to game the study, that is, to intentionally answer untruthfully in order to promote a particular agenda. For example, I might actually be willing to pay $5 to ride the subway, but I would rather pay less; so I report that I would never ride if the price were $5. This strategy is much more likely to occur to a respondent who is aware of the entire experimental design.

The hope is that Factorial Forecasting will be useful to people beyond the university. To facilitate that appreciation, it will be helpful to demystify the analysis and promulgate user-friendly software (Weiss, 2006). Functional measurement researchers, who already appreciate the incisive analytic capability of factorial designs, need to be sensitive to the pragmatic realities encountered when respondents are truly volunteers rather than campus draftees. Consultation between researchers and policy makers will be advantageous to both parties. A potentially querulous aspect of those conversations will be the divergent views of validity held by functional measurement researchers and policy makers. Policy makers will inevitably want to cling to the traditional criterion; is the prediction accurate? Can people project their future behavior? Could one predict as successfully by tossing a coin? The researcher needs to take the accuracy question seriously.

The simplest approach is to acknowledge the limitation that one simply cannot know about accuracy in a timely manner. Evidence can only be amassed after the policy has been implemented. And of course, it is impossible to tell whether a policy that was not implemented (because Factorial Forecasting predicted it would fail) would actually have been successful. External outcome validity (Anderson, 2001) can never be more than partially demonstrated.

It is not always easy for people to predict their future desires, particularly when emotions are evoked (Kahneman & Snell, 1992). Even the most earnest of respondents may fail to foresee future behavior correctly. For instance, the degree of control imposed by an addiction might be underestimated. Future behavior may also be constrained via legislative fiat. The effect of possible penalties for noncompliance can be explored within the experimental design. In a generally law-abiding society, the success of a policy that *allows* for behavioral change is perhaps less predictable than that of a policy that *requires* behavioral change.

An inevitable difficulty in forecasting is that the world may change between prediction and verification. The longer the time span before the policy can be put in place, the greater the chance for disruption. Some implementations, such as a proposed subway network, may not achieve full impact for many years. Changes may be evolutionary, as in the case of fuel prices or taxes that might go up or down. The analyst may allow respondents to consider such changes by incorporating various possibilities into the scenarios. Revolutionary changes may also occur. A technological breakthrough, such as the television or cell phone, that radically transforms behavior patterns might disrupt prediction if it happened to occur between research and implementation. Projections regarding subway usage could be dramatically wrong if, for example, an inexpensive personal hovercraft were to become available. The impact of a revolutionary change can only be studied if the analyst has sufficient foreknowledge to build it into the design.

This challenge for those who would predict has been elucidated in another context by Meehl (1954), who observed how a broken leg might override a therapist's, as well as the patient's, ability to predict when a depressed patient will resume normal activities. While an individual's unforeseen event would have little effect on the average in a moderately-sized sample, an innovation can affect behavior at the societal level. Meehl made the point that a broken leg, while unpredictable, is not a fudge factor but an easily observed phenomenon. The situations in which prophecy will fail should be identifiable once they arise.

Although personal prediction will not always be accurate, I contend that no other method for forecasting would work any better. To be sure, there are professionals who assert their expertise in predicting public behavior. However, the empirical literature is unsupportive of such claims (Armstrong, 1991; Tetlock, 2005). Factorial Forecasting does not ask anyone to predict how others will react to a new situation. Respondents are asked only how they themselves will act. The basic premise of the method is that I am the person best placed to predict what I will do.

Currently, laypersons provide predictions when public officials conduct town meetings or enlist focus groups. These popular methods also incorporate guesses regarding the behavior of others. Although they appear to provide the advantage of averaging opinions, potential weaknesses include biased sampling and the possibility that the group's views may be distorted by especially vocal members. In addition, the group setting may not foster due consideration of the full range of possibilities. An option that might have been acceptable in isolation is likely to be summarily rejected once the group settles upon the best choice.

A more philosophical phrasing of the validity issue frames the discussion in terms of correspondence and coherence theories of truth. Hammond (1996) observed that these different approaches lead to misunderstandings about validity. The correspondence approach would be to validate by showing that projected action does match future observed action. There have been a few laboratory studies that have observed matches between projections within a sample and behavior by the larger group (de Kort, McCalley, & Midden, 2008; Van Vogt & Samuelson, 1999), but I am not aware of evidence in the policy arena. Marketers have used this approach with some success (Louviere, 1988), comparing projected and eventual sales. However, the limitations on prediction accuracy discussed above convince me that the correspondence view of validity is impractical for policy questions. Still, if a forecast says that a policy will be successful and time proves that prediction incorrect, the proposer is likely to forego Factorial Forecasting in the future.

The coherence approach involves examining whether observations are consistent with a theory (Dunwoody, 2009). In functional measurement research, the test of whether an algebraic model describes the data exemplifies the use of a coherence criterion. In the forecasting context, we can look for relations among the predictions. As a subset of functional measurement, Factorial Forecasting looks to coherence as its validational cornerstone. But because the goal is prediction rather than model verification, the evaluation of coherence may be less rigorous than is customary in functional measurement research.

In Factorial Forecasting, support for a model is informative but need not be crucial. Also, a core concept in functional measurement, scaling of subjective values, is de-emphasized in this methodology. When the responses are numerical and a hypothesized model is consistent with the projected actions, scale values become available; they might well be of interest. But coherent data can guide a policy decision without knowledge of subjective values. For example, the hypothetical results in Figure 3 do not support a simple algebraic model, yet they still provide important guidance for the policy maker. However, I am not suggesting that responses should merely be accepted at face value. If the results had said that less convenience would generated more ridership, or that stations 10 miles from home generated more ridership than either 5 miles or 15 miles from home, then I would deem the results invalid. The coherence criterion cannot provide a guarantee that the projections will predict ridership, but coherent data do provide reassurance that the judgmental task is being taken seriously and the factors are influencing the respondents in a reasonable manner. Orderly data that exhibit main effects provide partial evidence of what Anderson (2001) has referred to as process validity.

Factorial Forecasting has the advantage of systematic variation of scenarios and random assignment of participants to scenarios. Any reliable differences between responses to different scenarios can be attributed to the manipulation. Thus, Factorial Forecasting is poised to profit from the internal validity of a true experiment, a *sine qua non* of science and applied research, such as medicine (in the form of randomized clinical trials).

I admit to some sympathy if a policy maker considers this empirical approach distasteful. One might adopt the position that the role of the policy maker is to do what is right regardless of how people will react; that is, to lead rather than to follow. Whether that position is ultimately seen as principled or arrogant will likely depend on whether the policy proves successful.

Policy makers may also have reservations about the practicality of engaging in factorial forecasting prior to implementation. Research not only incurs financial expense, but also requires several steps: design, recruitment, data collection, and analysis. There may be a concern that the policy window (Kingdon, 2010), a time-limited opportunity for implementation, will close before the results are available. On the other hand, failure to do the research may be far more costly over the long run, if the intuitions that inspired the new proposal turn out to be inconsistent with what people will do when the policy is in place. And if Factorial Forecasting does prove useful, it is likely that streamlined methods will become available, just as has occurred in the political polling industry.

I emphasize that the proposal is not to elicit values from the citizenry (Baron, 1997; Knetsch, 2002). Certainly, leaders ought to, and do, care about public attitudes (Weiss & Tschirhart, 1994); but attitudes contribute to rather than determine projected action (Weiss, John, Rosoff, Shavit, & Rosenboim, 2012). I contend that scenario research could have predicted the failure of Prohibition or of required busing to achieve racial integration in Los Angeles, just as it can predict whether a bike path will be used. All too often, policy makers rely upon the maxim espoused in the film Field of Dreams: "If you build it, they will come." History has not always supported that reliance.

## REFERENCES


Anderson, N. H. (1971). An exchange on functional and conjoint measurement: Comment. *Psychological Review, 78*, 457.

Anderson, N. H. (1996). *A functional theory of cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, N. H. (2001). *Empirical direction in design and analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, N. H., Sawyers, B. K., & Farkas, A. J. (1972). President paragraphs. *Behavior Research Methods and Instrumentation, 4*, 177-192.

Armstrong, J. S. (1991). Prediction of consumer behavior by experts and novices. *Journal of Consumer Research, 18*, 251-256.

Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin, 122*, 72-88.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods, 4,* 243-249.

Brengman, M., Wauters, B., Macharis, C., & Mairesse, O. (2010). Functional effectiveness of threat appeals in exercise promotion messages. *Psicológica, 31*, 577-604.

de Kort, Y. A. W., McCalley, L. T., & Midden, C. J. H. (2008). Persuasive trash cans: Activation of littering norms by design. *Environment and Behavior, 40*, 870-891.

Dunwoody, P. T. (2009). Theories of truth as criteria for judgment and decision making. *Judgment and Decision Making, 4*, 116-125.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self or social psychology? *Journal of Personality and Social Psychology, 79*, 861-875.

Gardiner, P.C., & Edwards, W. (1975). Public values: Multiattribute utility measurement for social decision making. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 1–37). New York: Academic Press.

Green, K. C., & Armstrong, J. S. (2011). Role thinking: Standing in other people's shoes to forecast decisions in conflicts. *International Journal of Forecasting, 27*, 69-80.

Green, P., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 1*, 61-68.

Green, P., & Wind, Y. (1971). *Multiattribute decisions in marketing: A measurement approach.* Hinsdale, Il: Dryden Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Hofmans, J., & Theuns, P. (2010). Testing the impact of predefined and self-defined end anchors on the linearity of the category rating scale. *Quality and Quantity, 44*, 397-408.

Howe, E. S. (1991). Integration of mitigation, intention, and outcome damage information by students and circuit court judges. *Journal of Applied Social Psychology, 21*, 875-895.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: do people know what they will like? *Journal of Behavioral Decision Making, 5*, 187–200.

Kingdon, J. L. (2010). *Agendas, alternatives, and public policy, update edition, with an epilogue on health care* (2nd ed.). New York: Longman.

Knetsch, J. L. (2002). Policy analysis and design with losses valued more than gains and varying rates of time preference. In R. Gowda & J. C. Fox (Eds.), *Judgments, decisions, and public policy* (pp. 91-115). Cambridge, UK: Cambridge University Press.

Krantz, D. H., & Tversky, A. (1971). An exchange on functional and conjoint measurement: Reply. *Psychological Review, 78*, 458.

Louviere, J. J. (1988). Conjoint analysis modeling of stated preferences: a review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy, 22*, 93-119.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and application*. New York: Cambridge University Press.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: a new scale type of fundamental measurement. *Journal of Mathematical Psychology*, *1*, 1–27.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist, 24*, 1063-1075.

Rundall, C. S., & Weiss, D. J. (1994). Nurses' fear of contagion: A functional measurement analysis. *Medical Decision Making, 14*, 40-46.

Rundall, C. S., & Weiss, D. J. (1998). Patients' anticipated compliance: A functional measurement analysis. *Psychology, Health, & Medicine, 3*, 261-274.

Shanteau, J., Pringle, L. R., & Andrews, J. A. (2007). Why functional measurement is (still) better than conjoint measurement: judgement of numerosity by children and adolescents. *Teorie & Modelli, 12*, 199-210.

Tetlock, P. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

Van Vogt, M., & Samuelson, C. D. (1999). The impact of personal metering in the management of a natural resource crisis: A social dilemma analysis. *Personality and Social Psychology Bulletin, 25*, 735-750.

Weikel, D. (2010, July 23). L.A. officials to mark 20th anniversary of Metro Rail system. *Los Angeles Times*. Retrieved September 19, 2011 from http://articles.latimes.com/2010/jul/23/local/la-me-rail-anniversary-20100723.

Weiss, D. J. (2006). *Analysis of variance and functional measurement: A practical guide*. New York: Oxford University Press.

Weiss, D. J. (2010). Using nominal data to examine information integration. *Psicológica, 31*, 441-459.

Weiss, D. J., John, R. S., Rosoff, H., Shavit, T., & Rosenboim, M. (2012). Fear of flying: Emotional and behavioral responses to escalating terrorist threat. Manuscript submitted for publication.

Weiss, J. A., & Tschirhart, M. (1994). Public information campaigns as policy instruments. *Journal of Policy Analysis and Management, 13*, 82-119.

Weiss, J. W., & Weiss, D. J. (2002). Recruiting Asian-American adolescents for behavioral surveys. *Journal of Child and Family Studies, 11*, 143-149.

(Manuscript received: 27 September 2011; accepted: 27 February 2012)