Psicológica (2010), 31, 441-459.

Using nominal data to examine information integration

David J. Weiss*

California State University, Los Angeles

Functional measurement studies typically collect numerical data in order to study judgment. The new Nanova (Nominal analysis of "variance") method allows for expansion of the paradigm to include the study of actual or projected behavior. In everyday life, people carry out actions that can be described using verbal labels, which are nominal data. Nanova is similar to analysis of variance in that significance statements assess the effect of experimentally manipulated factors. The way the methodology can extract cognitive strategies from behavioral actions is illustrated by considering a hypothetical burglar who attends to two safety features of the target homes under consideration. A real illustration is also presented, in which respondents reported both fear and projected actions in response to scenarios describing terrorist attacks. The emotional responses, reported as numbers, were analyzed with analysis of variance. The projected actions, reported nominally, were analyzed with the NANOVA computer program (Weiss, 2009). Two factors embedded in the scenarios, government announcement and public reaction, yielded similar effects on both kinds of response. Neither main effect was significant, nor were the anticipated effects of the variables obtained. With both response modes, the factors interacted significantly.

The hallmark of functional measurement methodology (Anderson, 1981) is its ability to reveal integrative processes by decomposing sets of judgments. Those judgments have traditionally been expressed as numerical

---

* Acknowledgments: This research was supported by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under grant number 2007-ST-061-000001. 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. I wish to thank Richard John, Heather Rosoff, Marcel Hovsepian, and Whitney Axley for their efforts on the study of responses to terrorism. I also wish to thank Professor John for valuable comments on an earlier version on the manuscript. Correspondence regarding this article, including requests for reprints, should be sent to David J. Weiss, 609 Colonial Circle, Fullerton CA. 92835 United States. Email: dweiss@calstatela.edu.

responses. Considering that action is often based on a judgment, we ought also to be able to gain insight into cognitive processes from how people choose to act. This paper introduces a new variant of functional measurement that decomposes nominal responses. Because the new method does not accomplish the simultaneous validation of model and response scale – there is no response scale with nominal data – it is not true functional measurement. Accordingly, I have borrowed another of Anderson's labels for the methodology he pioneered in the title of this paper.

Actions and projected actions are most conveniently described using words rather than numbers. One may plan to go to class or the beach, to buy a particular product, to drive or to walk. Researchers can observe action in the wild or simulate it in the laboratory. When plans or subsequent actions are elicited as a function of controlled variables, a set of nominal data arises.

If that set is sufficiently large, the analyst can describe the data using probabilistic terminology. Typically, significance testing involves Chi-square or one of its modern descendants such as multinomial logit analysis. A drawback is that these analytic techniques are disrupted if some responses occur infrequently. I am unaware of any previous functional measurement study that has employed nominal responses.

THE NANOVA ANALYSIS

I have recently developed a method (Weiss, 2009) for analyzing nominal responses, Nanova, that is similar to analysis of variance in that significance statements assess the effect of experimentally manipulated factors. The NANOVA computer program for carrying out the analysis, available at http://www.davidjweiss.com/NANOVA.htm, uses the CALSTAT suite interface (Weiss, 2006) and allows up to four factors. The program handles independent groups and repeated measures designs, as well as mixed designs because it permits factors to be nested. Its principal limitation is that equal cell sizes are required[1].

Nanova can be used with small samples and accommodates disparities among the frequencies with which the various response alternatives are employed. The import of this flexibility is that the researcher is free to choose whether to constrain the response options or to allow open-ended responses. There is a potential advantage unique to gathering unconstrained nominal responses: the subject may think of something the researcher never envisioned.

In the Nanova method, a variance-like quantity is computed by comparing pairs of responses. If two nominal responses do not match, then they vary, so the pair generates variance. In contrast, if the responses match, then they do not vary and the pair contributes nothing to the variance. This scheme is related to one independently proposed by Gini (1939), who by cleverly choosing the constants 0 for a match and 1 for a non-match, was able to define a true sum of squares for nominal data (because $0^2 = 0$ and $1^2 = 1$)[2].

The structure of the design determines the number of potential matches associated with each source. The data are parsed to find the proportion of matches that occurred. Those proportions are combined to yield *N*-ratios in the same way that mean squares are combined to yield *F*-ratios. The *N*-ratios are the test statistics whose significance needs to be assessed[3].

In the Nanova framework, significance questions are addressed using randomization tests (Edgington & Onghena, 2007). That method avoids assumptions about the distribution of the responses. The *p*-value reported by a test is the proportion of times a random permutation of the obtained responses yields an *N*-ratio greater than the *N*-ratio observed in the data.

The heart of the new method is the orthogonal assignment of pairs to sources. In a factorial design that generates N responses, each of the ${}_NC_2$ pairs of responses is uniquely assigned to a specific source. The assignment is a structural property of the design and is independent of the data. Just as degrees of freedom are allocated according to the design structure, so are pairs of potential matches.

In Table 1, I illustrate the assignment for a 3-factor, 2x2x2 design. With eight scores, there are ${}_8C_2$ (=28) pairwise comparisons that can be made. Each pair is allocated to a specific source, namely the one for which the indices for the two cells disagree. The number of comparisons is linked to *df*. The general rule is that the number of comparisons linked to a source is the product of the source's *df* and one-half the total number of scores. In this example, each of the seven sources has 1 *df* and accordingly has 4 associated comparisons.

Table 1. Assignment of the $_8C_2$ (= 28) pairs to sources for a 2x2x2 factorial design.

Cell indices	Compared to	Cell indices	Source
1 1 1		1 1 2	C
1 1 1		1 2 1	B
1 1 1		1 2 2	BC
1 1 1		2 1 1	A
1 1 1		2 1 2	AC
1 1 1		2 2 1	AB
1 1 1		2 2 2	ABC
1 1 2		1 2 1	BC
1 1 2		1 2 2	B
1 1 2		2 1 1	AC
1 1 2		2 1 2	A
1 1 2		2 2 1	ABC
1 1 2		2 2 2	AB
1 2 1		1 2 2	C
1 2 1		2 1 1	AB
1 2 1		2 1 2	ABC
1 2 1		2 2 1	A
1 2 1		2 2 2	AC
1 2 2		2 1 1	ABC
1 2 2		2 1 2	AB
1 2 2		2 2 1	AC
1 2 2		2 2 2	A
2 1 1		2 1 2	C
2 1 1		2 2 1	B
2 1 1		2 2 2	BC
2 1 2		2 2 1	BC
2 1 2		2 2 2	B
2 2 1		2 2 2	C

The design structure dictates how to set up the *N*-ratios. If the example were a repeated measures design with, say, C as the "Subjects" factor, then the *N*-ratios would be constructed by comparing the proportion of obtained to potential matches for A, B, and AB to those for AC, BC, and ABC respectively. If instead the design were independent groups with C the "Replicates" factor, then the Within Groups proportion of matches would be constructed by pooling obtained matches and potential matches for the AC, BC, and ABC sources. The pooled term would incorporate 12 comparisons, and would serve as the denominator of the *N*-ratios involving A, B, and BC.

The null hypotheses tested with nominal data are necessarily different from those used with numerical data. A factor does not affect behavior when responses to its various levels are the same. Accordingly, the obtained proportion of *non-matches* for the comparisons associated with that factor ought to be small if the null hypothesis regarding that main effect is true.

Matches relevant to the interaction between two factors are those for which the corresponding cell indices for the pair both differ (e.g., 11 vs 22, 21 vs 12, 31 vs 12, etc.). The more of those matches that occur, the greater the interaction. If the null hypothesis of no interaction is true, the obtained proportion of *matches* for the comparisons associated with the interaction should be small.

COGNITIVE ALGEBRA WITH NOMINAL RESPONSES

Functional measurement theorists seek to understand the process by which a person integrates information to produce a response. That entails determining whether the pattern of observed responses is consistent with a hypothesized algebraic model. The ambition does not change when nominal responses are involved, but the reduction in the amount of information contained within the data limits the conclusions that can be drawn. Distances between responses are not known. Accordingly, graphical representation is not possible, so one cannot investigate the parallelism characteristic of an additive model or the linear fan pattern that suggests a multiplicative process. Since there are no means, scales based on marginal means cannot be constructed.

These limitations notwithstanding, a nominal analysis can shed light on the integration process. I will support this claim first with a light-hearted illustration using a distorted version of a serious study carried out by Garcia-Retamero and Dhami (2009), then report some real data from a study of responses to terrorism threat.

Imagine that we are trying to learn how burglars view features of potential target homes. Our attention is concentrated on two safety features a home might have: a burglar alarm and a loud dog. Suppose we get some cooperative expert burglars who will helpfully tell us how they would react to a home (amazingly, Garcia-Retamero and Dhami were able to do this) characterized by the presence or absence of those features. I will ignore the practical issues involved in doing such a study and present five patterns of responses that a burglar might evince. The main point of the example is that when a factorial design is employed, simple responses expressing behavioral intention – Yes, I would rob that home, or No, I would not rob that home – are sufficient to reveal the operation of those two factors. It is worth emphasizing that we need all four cells of the 2x2 design to make inferences about the burglar's strategy.

Consider the patterns of responses shown in Tables 2-6. For each pattern, a distinctive strategy can be verbalized, one that corresponds to a specific Nanova result. To test for statistical significance, the researcher would need to generate an error term in one of the usual ways.

Table 2. Robbery pattern for alarm-phobic burglar.

	Alarm	No alarm
Dog	Don't rob	Rob
No dog	Don't rob	Rob

Table 3. Robbery pattern for dog-phobic burglar.

	Alarm	No alarm
Dog	Don't rob	Don't rob
No dog	Rob	Rob

The alarm-phobic burglar exemplified by Table 2 generates only a main effect for the alarm factor; his dog-phobic counterpart shown in Table 3 generates only a main effect for the dog factor. The report of a significant main effect does not convey the direction of the effect; one must look at the data to see which level of the factor is associated with which response.

Table 4. Robbery pattern for indiscriminate burglar.

	Alarm	No alarm
Dog	Rob	Rob
No dog	Rob	Rob

Table 5. Robbery pattern for cautious burglar.

	Alarm	No alarm
Dog	Don't rob	Don't rob
No dog	Don't rob	Rob

The indiscriminate burglar in Table 4, who robs all homes, generates no effects. His behavior is independent of the manipulated factors. In contrast, the cautious burglar in Table 5 generates all three effects (main effects for alarm, dog, and their interaction). This careful professional might really think additively, reasoning that a home with two protective features is a distinctly less appealing target than a home with only one. However, the response options in this case are insensitive to the degree of the target's unattractiveness. This example highlights one of the limitations of studying pure behavior. Even if the two features do add at the cognitive level, we cannot observe the more extreme internal reaction. We would be able to see an additive pattern only if we allowed for other responses, such as an emotional scream that might erupt only when both safety features were present.

Table 6. Robbery pattern for Topkapi burglar.

	Alarm	No alarm
Dog	Rob	Don't rob
No dog	Don't rob	Rob

Perhaps the most interesting burglar is shown in Table 6, whom I named Topkapi after a classic film in which a lazy small-time crook is induced to participate in robbing a well-guarded palace that contains great treasures. In this case it is only the interaction that stands out.

REAL DATA

Because the avowed objective of terrorism is to disrupt lives, my research team has focused on two kinds of responses to the threat. Terrorism can inspire fear and motivate behavioral changes to reduce exposure. Our paradigm calls for subjects to read scenarios describing nearby attacks. We then ask for reports of emotional impact and projected action. In the study I describe here, the scenarios described a MANPAD attack (Okpara & Bier, 2008; von Winterfeldt & O'Sullivan, 2006). We told respondents (accurately) that "a MANPAD is a shoulder-fired missile that can bring down a low flying aircraft. About the size of a bag of golf clubs, it can be carried by one man and concealed in the trunk of a car. Many were distributed to Afghans fighting the Russians during the 1980's. Several thousand remain unaccounted for."

We then added an acknowledged fictional account of a MANPAD attack: "NBC News has reported that an international terrorist organization is planning to use Man-Portable-Air-Defense-Systems (MANPADS) to cripple the airline industry. Intelligence reports suggest the uncovered plot was in the advanced stages of planning. The terrorists had mapped out launch locations within a couple of miles of airports. These locations were believed to be selected because they allow shooters to simply drive to a good vantage point and avoid airport security altogether. Sources also confirm that an email was recently intercepted asserting that 300 MANPADS have been shipped from Iran to the US. There is currently no technology installed within aircrafts to counter the missiles."

Our respondents were undergraduates given access to a private web site on which the study was presented. All respondents saw the incident described above. In addition, we manipulated government announcement and public reaction to this event. Both factors had three levels. The levels of the manipulated factors were also conveyed as fictionalized news reports. The three levels of government announcement were:

Level 1, which we refer to as Plot Foiled: "Following the NBC News story, a Department of Homeland Security spokesperson said they have the situation under control. The ringleaders of the plot have been arrested, and almost all of the MANPADS en route to the U.S. have already been confiscated. He added that existing security measures led to the arrest of persons involved and urged the public to feel comfortable going about their daily business as usual."

Level 2, which we refer to as Evaluating: "Following the NBC News story, a Department of Homeland Security spokesperson said they are still looking into the nature of the threat and have some leads they are pursuing aggressively. He added that the Department is evaluating security in place at airports and asks that the public be extra vigilant in reporting any suspicious behavior to their local authorities."

Level 3, which we refer to as Serious Problem: "Following the NBC News story, a Department of Homeland Security spokesperson said that MANPADs represent a serious threat. He noted that at this time they have no strong leads regarding the ringleaders or the locations of the MANPADS. The Department has decided to close down the airport for the next three days to evaluate security in the surrounding vicinity."

The three levels of public reaction were:

Level 1, Flying rates increased: "CNN just reported that during the two weeks since the story broke, the major airlines announced a ticket price cut, and sales have increased by 20%."

Level 2, Flying rates steady: "CNN just reported that during the two weeks since the story broke, the major airlines announced a ticket price cut, and sales have remained steady."

Level 3, Flying rates decreased: "CNN just reported that during the two weeks since the story broke, the major airlines announced a ticket price cut, but sales have dropped by 40%, and 30% of scheduled passengers have been no-shows."

Twenty-four respondents each saw one of the nine combinations (Total N = 3 x 3 x 24 = 216). All respondents answered the same questions. Typical of our questions exploring an emotional response was the fear item: "How fearful would you be after learning of this terrorist plot?", to which the response was made on a 10-point scale where 10 represents the highest degree of fear. This question gives rise to a traditional functional measurement analysis using numerical responses.

We anticipated the fear responses might look something like Figure 1, an additive pattern. We expected that when the plot was foiled, people would feel safe because their government knows how to protect them. When the problem is seen as a serious one, people would feel frightened. We also expected people to be sensitive to the social cue. When flying rates drop, people think their neighbors are frightened. Accordingly, they also should be frightened.

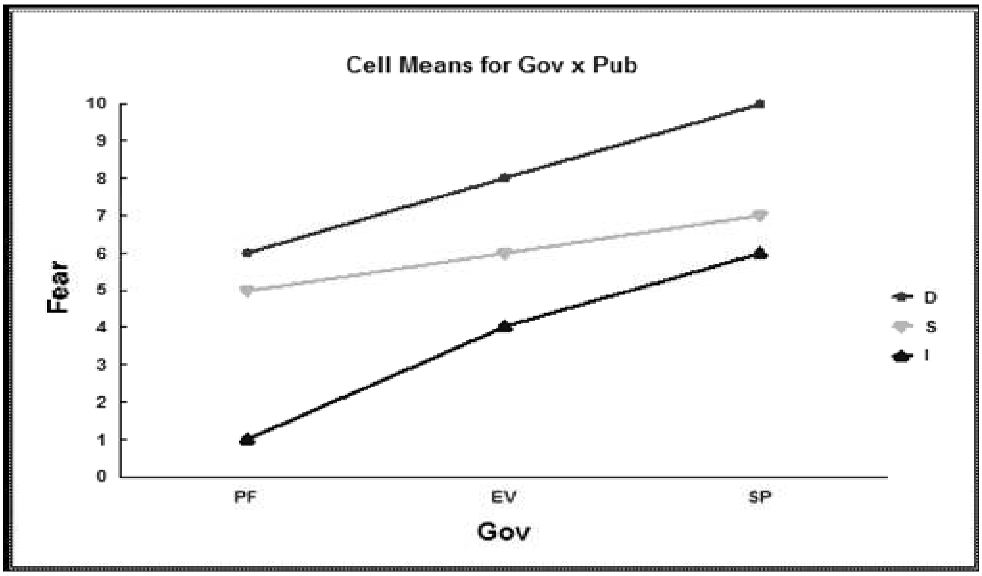

Figure 1. Anticipated results for fear as a function of government announcement (Plot Foiled, Evaluating, Serious Problem) and public reaction (Decrease, Steady, Increase).

However, the classical parallel plot we anticipated did not obtain. The actual responses, shown in Figure 2, evince only moderate fear, and our manipulated variables did not influence the behavior in a way that I can make sense of. The analysis of variance results shown in Table 7 confirm the graphic results, in that only the interaction proved significant. These results were not unique to the fear response; we asked similar questions about worry ("To what degree would you be worried after learning of this terrorist plot?") and risk ("To what degree would you feel that you were at risk after learning of this terrorist plot?"), and obtained similar results. The graph for the worry response is shown in Figure 3. The pattern is almost identical, although the cell means were all about one point higher than for fear.

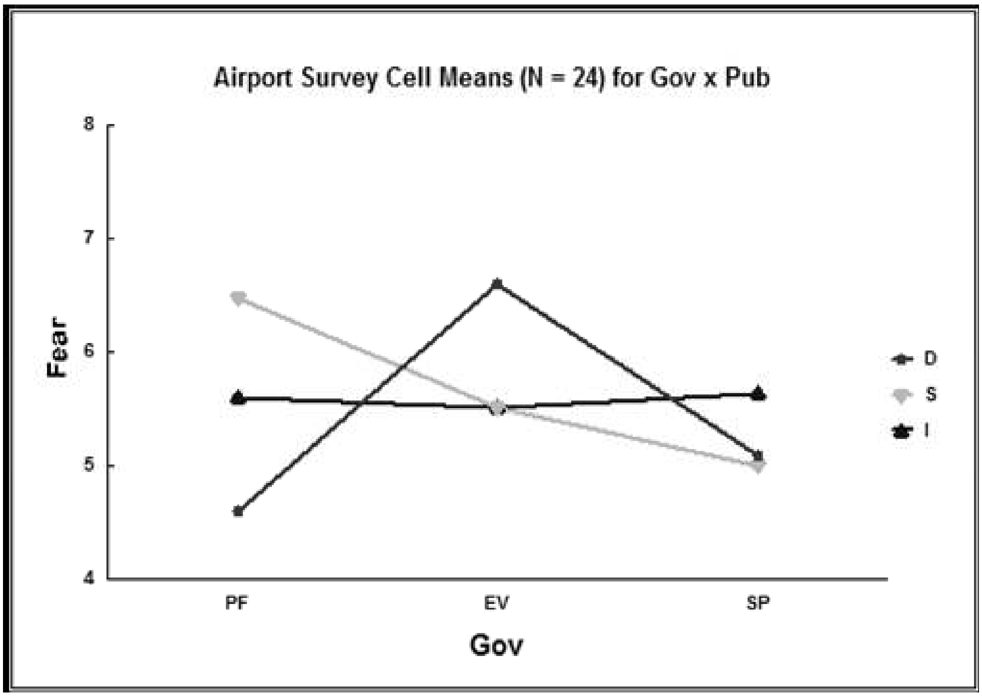

Figure 2. Obtained results for fear (10-point scale) as a function of government announcement (Plot Foiled, Evaluating, Serious Problem) and public reaction (Decrease, Steady, Increase).

Table 7. Analysis of variance for fear response.

Source	*df*	MS	*F*
Gov	2	7.03	1.36
Pub	2	1.03	<1
GxP	4	16.12	3.13*
Error	207	1066.92	

*$p < .05$.

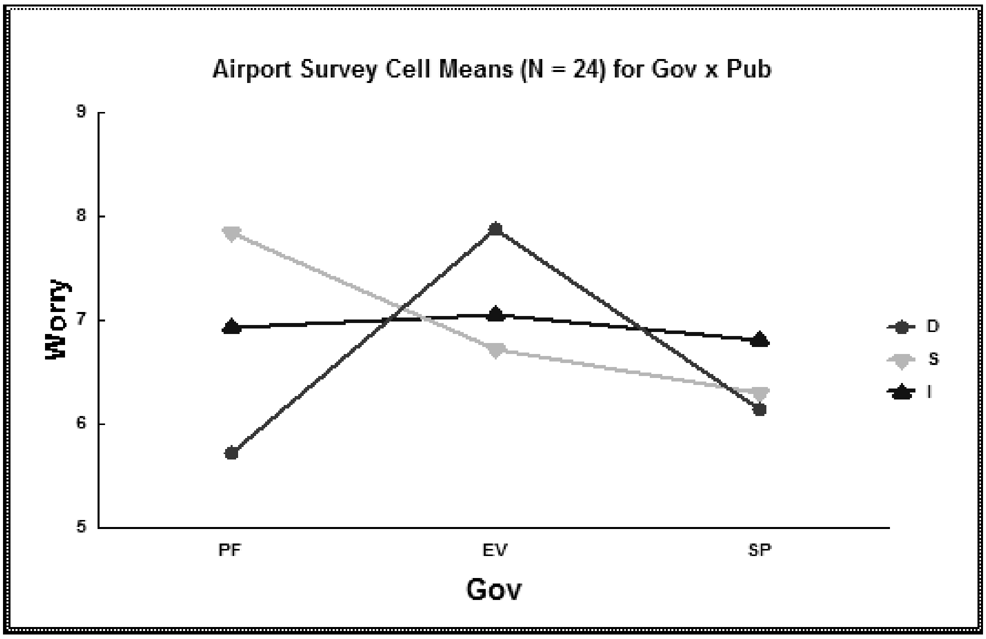

Figure 3. Obtained results for worry (10-point scale) as a function of government announcement (Plot Foiled, Evaluating, Serious Problem) and public reaction (Decrease, Steady, Increase).

These disappointing results were obtained with a between-subjects (also called independent groups) design, which is not the norm in functional measurement studies. We chose this design advisedly, being concerned that a repeated-measures design might yield artifactual model support. The issue is controversial and will be addressed in general terms in the discussion section.

Our choice was made to fit the research context. We were interested in emotional responses to a dramatic event. Exposing the subject to a series of such events, with the government response and public reaction visibly changing, would change the nature of the task. Presenting a series is likely to inspire logical comparisons, perhaps even inducing the subject to generate a hypothesis about the design. Orderly data might well be produced, as the subject attempted to respond to the stimulus variation in accord with that hypothesis; but those data would no longer express feelings.

In the present case, I do not ascribe the failure of the data to support the additive model to a lack of power. We ran a large number of subjects per cell. In fact, for all three of the numerical responses, the interaction was significant although the main effects were not, thus illustrating that there was sufficient power to find effects.

The simplest explanation is that the variables we manipulated were ineffective. In particular, Americans may have learned to ignore government announcements. The Department of Homeland Security "Orange" threat level has been in force almost continuously since 9/11, although no events have occurred on American soil; these are optimal conditions for a "cry wolf" effect (Breznitz, 1984) to occur. It is also possible that the levels were interpreted differently by individuals. For example, to some respondents, foiling the plot might constitute evidence the threat is real, whereas evaluating allows for the belief there is no true threat. An inevitable characteristic of the between-subjects design is that one cannot untangle potential variation in connotative meaning. Another possibility is that the terrorist plot evoked only moderate fear, perhaps because our respondents were children during the 9/11 attack. Attempts to manipulate their fear thereby had little to work with.

We did employ a within-subject manipulation when we collected the nominal responses. In addition to reporting fear as described above, respondents also were asked to imagine 3 forthcoming trips in which they were scheduled to attend an event 1500 miles away within one month of the terrorist incident. We attempted to manipulate the importance of the trip. The three levels of importance were (1) best friend's wedding (2) job interview with wonderful prospects (3) long-awaited vacation with friends and family. Structurally, this is a mixed design, in which subjects were nested under government announcement and public response and crossed with trip importance. We were not concerned about artifactual model support in this case because from an individual subject's perspective, only trip importance was varied.

The response was open-ended, so any words could be entered. The simplest way to deal with nominal responses is to analyze the words as the subject typed them. However, accepting responses literally runs the risk of missing the meaning. People use alternative phrasings of the same idea, they make typing errors – and in neither of those cases does it seem appropriate to regard discrepancies as non-matches. We had human experts (graduate students) preprocess the data by judging the intention. We anticipated that most of the responses would come from a small set of possibilities, including staying home, driving (we chose the distance of 1500 miles to make that option unattractive), or attempting to postpone or switch the location. In fact, the judgment of our experts was hardly taxed, because the most common response by far was to fly as planned.

The Nanova results are shown in Table 8. For our purposes what matters are the *p*-values, which are interpreted analogously to those in analysis of variance. A *p*-value smaller than .05 means there were different responses to the various levels of the source. Here, what we see is that the manipulated variables on which we focused previously yielded similar results. The interaction was significant, but the main effects were not. The importance manipulation also had an impact only through its interactions.

Table 8. Nominal analysis of variance for projected action.

Source	*df*	Potential Matches	NP	*N*-ratio	*p*
Gov	2	648	.367	.61	1.0
Pub	2	648	.392	.66	1.0
GxP	4	1296	.635	1.06	<.001
Error	207	67068	.598		
Imp	2	648	.360	.64	1.0
GxI	4	1296	.558	1.00	<.001
PxI	4	1296	.547	.98	<.001
GxPxI	8	2592	.564	1.03	<.001
Error	414	134136	.558		

Additionally, we examined a quantitative counterpart of the projected action, asking for the likelihood of flying on a major airline to the destination. Because flying to the destination is akin to the behavioral option of going on the scheduled trip, we might expect likelihood of flying to address the same cognition. That expectation can be examined by carrying out a traditional functional measurement analysis on the numerical likelihood responses.

In a global sense, the results for the two types of data were similar, in that most reported likelihoods were high just as most respondents stated they would take the trip as planned. However, when we examine the results in detail, we see discrepancies. In contrast to the behavioral responses, the likelihood judgments produced a significant effect of trip importance in the expected direction (lower likelihood for the vacation). Also, government announcement had a significant, albeit small, effect, while none of the interactions was significant. The analysis of variance for this mixed design is shown in Table 9. The pattern of results can be seen in Figure 4. These data look more like what functional measurement researchers are used to seeing, except that one of the manipulated factors, public reaction, did not yield a significant effect.

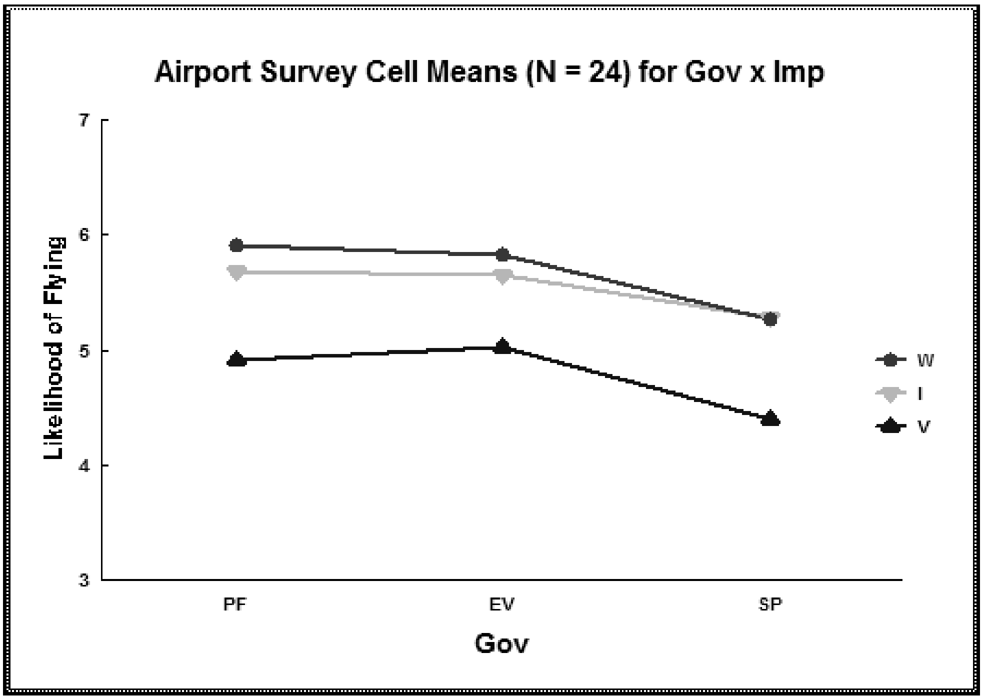

Figure 4. Obtained results for likelihood of flying (7-point scale) as a function of government announcement and trip importance (Wedding, Interview, Vacation).

Table 9. Analysis of variance for likelihood of flying.

Source	*df*	MS	*F*
Pub	2	1.44	.25
Gov	2	19.36	3.42*
PxG	4	.20	.03
Error	207	5.66	
Imp	2	49.80	55.56*
PxI	4	.76	.85
GxI	4	.56	.63
PxGxI	8	.75	.84
Error	414	.89	

*$p < .05$.

I do not believe these differences in outcomes reflect a native difference in statistical power between the analytic methods. Rather, they occur because the limited number of plausible behavioral options available compresses a range of high likelihoods into one reported action (fly as planned). In this instance, the projected behavior itself is insensitive to cognitive variation. Relatively small differences in likelihood as a function of trip importance can be detected by analysis of variance, but projected behavior is the same. In addition, the lower likelihoods for the vacation trip correspond to a variety of possible actions. If the actions chosen by those who elect not to fly vary, there will no consistent pattern for Nanova to find, and no main effect.

The interactions that appeared in the nominal responses confirm that trip importance, government response, and public reaction do not contribute independently to whether one will fly as planned. The nonsignificance of the main effects suggests that the interactions are of the crossover type illustrated in Table 6.

DISCUSSION

The empirical debut of the Nanova technique did not yield easily summarized results. The analysis looks more convincing when I get to make up the data, as in Weiss (2009). Here, the moderate levels of reported fear suggest that the respondents were not very terrorized by our scenarios. In future work, we will increase the level of gore and make the presentation more vivid. Another tack is to employ respondents whose lives have been touched by terrorism, so that the threat is personally relevant.

The most important reason to add Nanova to the functional measurement researcher's toolbox is that judgment and action are subject to different constraints. A numerical response can take on any value within a range defined by the researcher. However, the number of plausible actions or projected actions is often limited by practical considerations. What could our respondents do about the scheduled trip? They could hardly take up arms against the terrorists. It's hard to conceive of sensible actions other than to fly as planned, stay home, or use alternate transportation. Attempting to reschedule the vacation is possible, but not the wedding or the interview. It is worth reiterating that the difference between the results using the likelihoods as opposed to the behavioral responses suggests something of psychological interest. The different levels of safety inherent in the three government announcements affect likelihoods in the expected way, but their impact on projected action is less clear cut. When one thinks in terms of what actions might be taken, subtle shades of intention cannot be expressed.

There may also be other reasons for inconsistencies between attitude and action. For example, I might judge lobster to be preferable to pizza, but in the restaurant I order pizza because of the price differential. The inconsistency arises because cost, a dimension that is not important to a purely affective preference response, is relevant to choice. In the terrorism study, action may well have been affected by motives other than fear. If a respondent viewed the ticket as non-refundable (this was not addressed in the scenarios), sunk cost might come into play and promote going on the trip despite heightened fear. The distinction between attitude and action has been of interest to researchers for a long time (LaPiere, 1934), but cognitive modelers in general and functional measurement researchers in particular have confined their attention to what people say rather than to what they do.

Admittedly, studying projected action is not the same as studying real action. People may not behave in accord with their stated projections. This divide presents an exciting challenge to the researcher. Can we employ factorial designs in contexts where the behavior is both meaningful to the subject and observable by the investigator?

Studying action also raises methodological challenges. Although the mathematics of power analysis have not yet been worked out for Nanova, the general picture appears to be similar to that for analysis of variance. The more observations, the better; and variability between people is inevitably greater than variability within people.

Of concern is a possible bias in the Nanova procedure that may have contributed to the repeatedly obtained result of nonsignificant main effects and significant interaction. In data sets where there are many matches, interactions are likely to achieve significance. This occurs because nominal interactions arise when responses across the relevant cells *match*. In contrast, nominal main effects arise when responses across the relevant cells *do not match*. The manner in which this distinction is interwoven with the distribution of responses remains to be investigated[4]. In defense of the nominal analysis, I reiterate that the same pattern, nonsignificant main effects along with significant interaction, appeared with both the numerical (fear, worry, risk) and nominal (action) response modes.

The use of a between-subject design violates tradition. Birnbaum (1999) has harshly criticized between-subject designs, having demonstrated that idiosyncratic use of the response scale can generate absurd relations among cell means. A less dramatic danger is that systematic differences among people inflate the error term in a between-subjects analysis, resulting in low power.

Such considerations often underlie the strong industry preference for single-subject and repeated-measures designs (Weiss, 2006). The potential drawback to those solutions is that actions tend to be easily remembered, thereby violating the critical experimental assumption that responses are independent. When a person is exposed to the various stimuli in a factorial design, contrast effects are a threat to validity. Even worse, when a stimulus is repeated, if the respondent merely recalls and repeats the previous response, the apparent gain in power is spurious.

The danger of artifactual model support is perhaps not severe when the stimuli are difficult for the subject to identify. In psychophysical studies, for instance, the stimuli are hard for the subject to label and many are virtually indistinguishable. Furthermore, if the stimulus design is large, the memory burden is sufficiently challenging that subjects cannot recall previous responses. However, in the terrorism study, not only were the stimuli particularly vivid, but also a large design would be ecologically unrealistic and expected to produce emotional habituation.

The researcher can attempt to mitigate these problems, perhaps by altering irrelevant aspects of the stimuli or by separating the trials over time. A time-honored tactic, one I relied upon in my early work on psychophysical judgment, is to make the stimuli so boring that the subject doesn't try to recall previous responses. The fallback position, when none of these solutions appears sufficient, is to employ an independent groups design. The downside of that resolution is that inter-individual variability then becomes a major concern. Nested group designs (Rundall & Weiss, 1998) may offer a compromise when they are feasible. Even better, it is possible in some situations to use nesting to experimentally extract the between-subjects contribution (Masin, Crestoni, & Fanton, 1988).

When effects are large, a within-subjects design can find them, as has been demonstrated in functional measurement studies reported by Howe (1991) and by Egu and Weiss (2003). And it is after all large main effects that we should wish to find. Grice (1966) drew the conclusion that between-subjects and within-subjects designs should be regarded as essentially different experiments. Functional measurement researchers are well aware of this procedural tension (Anderson, 2001).

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

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

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

Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Edgington, E. S., & Onghena, P. (2007). *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

Egu, C. L., & Weiss, D. J. (2003). The role of race and severity of abuse in teachers' recognition or reporting of child abuse. *Journal of Child and Family Studies, 12,* 465-474.

Garcia-Retamero, R., & Dhami, M. K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review, 16*, 163-169.

Gini, C. (1939). *Variabilità e concentrazione*. Vol. 1 di: *Memorie di metodologia statistica*. Milano: Giuffrè.

Grice, G. R. (1966). Dependence of empirical laws upon the source of experimental variation. *Psychological Bulletin, 66,* 488-498.

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.

LaPiere, R. T. (1934). Attitudes and actions. *Social Forces, 13*, 230-237.

Light, R. J., & Margolin, B. H. (1971). An analysis of variance for categorical data. *Journal of the American Statistical Association, 66*, 534-544.

Masin, S. C., Crestoni, L., & Fanton, V. (1988). Test of the absolute effect of size on heaviness. *Bulletin of the Psychonomic Society, 26*, 423-424.

Okpara, U., & Bier, V. M. (2008). Securing passenger aircraft from the threat of Man-Portable-Air Defense Systems (MANPADS). *Risk Analysis, 28*, 1583-1599.

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

von Winterfeldt, D., & O'Sullivan, T. M. (2006). Should we protect commercial airplanes against surface-to-air missile attacks by terrorists? *Decision Analysis, 3*, 63-75.

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

Weiss, D. J. (2009). Nominal analysis of "variance". *Behavior Research Methods, 41,* 901-908.

(Manuscript received: 28 May 2009; accepted: 24 September 2009)