Optimism bias and differential information use in supply chain forecasting

Robert Fildes

Dept. Management Science,

Lancaster University, LA1 4YX, UK

Email: R.Fildes@lancaster.ac.uk

Paul Goodwin*,

School of Management, University of Bath

Bath, BA2 7AY,UK

Tel: 44-1225-383594

Fax: 44-1225-386473

Email: mnspg@bath.ac.uk

August 2011

* Corresponding author

Acknowledgements

This research was supported by a grant from SAS and the International Institute of Forecasters.

Abstract

Companies tend to produce over optimistic forecasts of demand. Psychologists have suggested that optimism, at least in part, results from selective bias in processing information. When high demand is more desirable than low demand information favouring high demand may carry more weight than information that suggests the opposite. To test this, participants in an experiment used a prototypical forecasting support system to forecast the demand uplift resulting from sales promotion campaigns. One group was rewarded if demand exceeded an uplift of 80%. All of the participants were also supplied with information some of which supported an uplift of greater than 80% and some of which suggested that the uplift would fall below this value. This enabled an assessment to be made of the extent to which those who were rewarded if the uplift exceeded 80% paid more attention to the positive information than those who were not rewarded. While the experiment provided strong evidence of desirability bias there was no evidence that the group receiving rewards for sales exceeding the threshold made greater use of positive reasons. Possible explanations for this are discussed, together with the implications for forecasting support system design.

Introduction

Demand forecasts in supply chain organizations guide decisions relating to sales, marketing, finance, distribution, purchasing and operations and their reliability can therefore be a crucial factor in how these organizations perform. These forecasts usually rely partly on software which uses statistical algorithms to produce estimates of future demand. However, forecasters often receive additional information that is not available to the computer system. This may relate to facts (e.g. about a customer's forthcoming sales promotion), estimates (e.g. the expected outcome of the promotion) or the trustworthiness or confidence that can be placed in these facts or estimates.

To take this information into account it is common practice to make judgmental revisions to the system forecasts. Indeed, many companies revise more than 50% of their system forecasts in an attempt to improve accuracy. The effective interpretation of the additional information received is therefore critical. However, psychologists have identified a number of cognitive biases in the way in which unaided people use information and advice. Consistent with this, our earlier research (Fildes et al, 2009) has found that forecasters in companies are often unable to make effective use of additional information and they find it difficult to integrate this information with the forecasts generated by the computer system. The resulting forecasts are often biased and inefficient and, in a large percentage of cases, the judgmental revisions have the effect of reducing accuracy. A particular problem that we identified is the tendency for adjustments to be over optimistic. We found that, when forecasters in three supply chain companies judgmentally adjusted their statistical forecasts upwards, it was common for these adjustments to be in the wrong direction (i.e. actual sales turned out to be lower than the statistical forecasts) or in the right direction but so large that forecast accuracy was reduced.

This paper reports on an experiment that was designed to investigate, in a laboratory setting, the extent to which this optimism resulted from the differential use of available information, with information favouring the desirable outcome of high sales receiving more attention than information suggesting lower sales. If this bias is evident it will be potentially valuable in guiding the design of forecasting support systems. For example, such systems may try to mitigate the bias by drawing attention to negative information or by providing guidance on how information should be used (Goodwin et al, 2011).

Desirability bias in forecasting

Our earlier research and the broader literature on information and advice usage have identified a number of potential reasons why additional information is not used effectively by forecasters in supply chain companies (Fildes et al, 2009).

1. Decisions to adjust statistical forecasts are often made unnecessarily because forecasters are seeing false patterns in the noise associated with time series. In particular, forecasters have a propensity to create explanations for these random movements, when evidence to support these explanations is absent.
2. Information and advice that favours optimistic forecasts of demand is weighted more heavily than information suggesting lower levels of demand, even if the 'positive' information is less reliable. The result is that demand forecasts tend to be too high
3. When adjusting statistical forecasts, forecasters often rely on their ability to recall analogous events (e.g. similar promotion campaigns) or other information and advice. This means that recent or salient events, or recently delivered information and advice, tend to

attract disproportionately high weightings when the forecaster has to determine what size of adjustment to make. In addition, the effect on demand of analogous events that occurred some time ago may be wrongly recalled by the forecaster.

4. Advice from human experts is regarded as more reliable than that from statistical methods (Onkal et al, 2009). Similarly, information that is delivered ‘in person’ may attract more attention than information which is delivered electronically.
5. Information and advice that conflicts with the forecaster’s initial views may be discounted. Similarly, conflicting indications between different advisors and different items of information may lead to the discounting of all available information and advice.
6. Forecasters may make inaccurate assessments of the reliability of information and advice by recalling isolated and unrepresentative past instances where the advice and information proved to be unreliable. Also, surprisingly few companies look back to examine the extent to which whether their judgmental adjustments have improved accuracy so that inaccurate beliefs on the reliability of different information sources and advisors may persist.

Of these, the second reason is of most interest in this paper. Apparent optimism can result from a number of causes. Where a forecast is subject to asymmetric loss (e.g. where stock-outs are more costly than over stocking) there may be a tendency too forecast too high. However, we regard such apparent optimism as resulting from a confusion of decision making with forecasting(Goodwin, 1996). Similarly, sales forecasts may be inflated for political reasons, rather than from a genuine (albeit mistaken) belief that sales will be high. In these cases we would not regard optimism as the cause of any resulting forecasting errors. Optimism occurs when the desire for an event to occur leads to a genuine belief that the event is more likely to occur than it really is. In this respect the term desirability bias is probably a more useful term than optimism bias.

Krizan and Windschitl (2007) have identified nine possible mechanisms that can lead to desirability bias (these mechanisms are not mutually exclusive and may interact). Of these, three appear to be potentially relevant to demand forecasting in supply chains.

1. *Valence priming*

Positive valence refers to the intrinsic attractiveness of an event (e.g. high sales). The positive valence of an event leads to enhanced activation of similarly valenced information For example, if you will win £1000 if company A, rather than B wins a contract, the positive valence of company A winning means that attention is paid more to its positive characteristics (long history, reliable reputation) than its negative (e.g. recent layoffs). Thus, in the case of demand forecasting, the desirability of high sales may activate attention to factors that suggest a rise in sales and so that negative factors that are present are neglected.

2. *Confirmation bias*

In this mechanism people frame hypotheses they consider according to what is desirable. For example, a sales manager might hypothesise: “a *rise* in sales will occur” and then he or she will search for evidence consistent with this hypothesis in order to confirm it. Note that confirmation bias is distinguished from valence priming in that it relates to the search for information, rather than the use of already-available information

3. *Differential scrutiny of information*

In this case evidence in favour of the desired outcome (e.g. high sales) is accepted at a relatively low quality threshold. Evidence against the desired outcome is more carefully scrutinised so that weaknesses in *this* evidence are more likely to be uncovered.

In this paper we will focus on the role of valence priming in driving desirability bias. However, there is a good reason why this might not apply. Krizan and Windschitl (2007) have also identified a negativity bias where a strong desire for an outcome may make information that is inconsistent with that outcome have a large affective impact on the forecast. In addition, it is known that in some circumstances negative information has more salience and hence greater impact than positive information (e.g. see Rozin & Royzman, 2001).

To our knowledge desirability bias, and in particular valence priming as a cause of the bias, has not been studied before in a realistic setting. Existing studies relate only to probability estimates or the prediction of which of two events will occur (e.g. whether a picture card or non-picture card will be drawn from a pack of cards). To remedy this serious gap in the literature we conducted an experiment which we describe next.

Details of experiment

Forty-six mainly management students at the Universities of Bath and Lancaster took part in the experiment. Their task was to assess the extent to which statistical one-month ahead demand forecasts produced by a grocery manufacturer needed to be adjusted to take account of forthcoming product promotions. The students were randomly assigned to one of two groups. The CONTROL group (n= 21) simply received a flat payment of £6 for taking part in the experiment. The "High sales desirable" (HSD) group (n=25) were told that they would receive 100 points reward for each product if the sales uplift resulting from the promotion exceeded 80%. This uplift is typical of promotion effects in practice and both groups were informed that this level of uplift reflected the average effect of promotions in this company. Each 100 points earned translated into a £1 payment.

The interface used in the experiment is a prototypical forecasting support system and a typical screen shot is shown in figure 1. It can be seen that, for each product, a series of the most recent 24 past monthly sales figures was displayed (this relatively short series length is common in practical forecasting situations), together with statistical forecasts. The sales figures displayed were simulated sales series which were designed to reflect a range of typical patterns observed in company demand forecasting. The noise associated with the series was normally distributed and half of the 24 series had a high noise standard deviation of 80 units and the other half a low standard deviation of 40 units. Mean sales levels were around 200 units. Within each of these groups, four series had no trend, four had a slight upward trend of +1% and four a slight downward trend of -1%. Each series displayed was disturbed in one month by the effect of a single past promotion. Across the series these promotions caused a mean uplift of 80% in the underlying demand. The statistical forecasts were produced by exponentially smoothing the past sales figures using a smoothing constant of 0.2.

Please insert figure 1 about here

Please insert figure 2 about here

The participants made a one-month ahead forecast for each of the 24 products which were presented in a random order, after which they answered an end-of-experiment questionnaire (see appendix). For each forecast four reasons were supplied to inform the forecaster's judgment. These related to the likely effects of the weather, market research results, the anticipated effectiveness of the campaign and details of the spending on the promotion. One to three of these reasons were positive (overall 50% positive). Overall, there were four sequences of reasons that were designed to counterbalance each other. Participants within each ‘reward’ treatment (CONTROL and HSD) were randomly assigned to receive one of these four sequences of reasons. The purpose of providing four sequences of reasons was to remove the effect of any biases that might emanate from preconceptions, for example about the reliability of weather forecasts or biases arising from the order of presentation of the reasons. Thus each sequence of reasons was balanced with a ‘mirror’ set so that, for every positive reason relating to the weather, for example, there was a sequence involving a negative weather-related reason. Two further sequences were created by reversing the ‘original’ and ‘mirror’ sequences.

Figure 2 displays a typical selection of reasons favouring a greater than 80% uplift (positive reasons) and those suggesting an uplift of less than 80% (negative reasons). For each product, participants were asked to select two main reasons, from the four provided, to indicate their rationale for making their judgmental adjustment to the statistical forecast. We hypothesised that those in the HSD group would (a) make larger upward adjustments to the statistical forecasts and (b) have a greater propensity to select positive reasons.

Results

As hypothesised, the results strongly supported a tendency for the HSD group to make larger upward adjustments than the CONTROL group. The mean percentage adjustment (i.e. the adjustment as a percentage of the statistical forecast) was +38.0% for the HSD group and +18.7% for the CONTROL group. Because the distribution of adjustments were non-normal a Mann-Whitney test was used to compare the median of the mean adjustments made by the participants in each group. The respective medians were +40.1% and 10% and the difference between these was highly significant ($p=0.0056$). A possible explanation for the difference is that participants in the HSD group were anchoring their adjustments on 80%, rather than manifesting desirability bias. The distributions of mean adjustments made by members of the two groups are displayed in figure 3 and provide no evidence at all for a tendency to anchor.

Please insert figure 3 about here

What use did the participants make of the provided reasons? In answer to the question: "The reasons had a direct influence of my own forecasts. 1. Strongly disagree…… 5. Strongly agree" the mean response of the HSD group was 3.9 while it was 4.0 for the CONTROL group. Although this difference was not significant the overall mean of 3.8 was significantly different from the 'non-committal' response of 3.0 ($t=8.57$ $p<0.0001$). This shows that participants in both groups <u>said</u> they were influenced by the reasons. Moreover, there were significant correlations of 0.31 for the HSD group and 0.31 for the CONTROL group ($p<0.0001$, one tail) between the number of positive reasons *available for selection* and the size of the percentage adjustment. The correlations between the number of positive reasons *selected* and the percentage adjustment were even greater (0.43 and 0.38 for the two groups respectively).

However, there was no evidence, contrary to our hypothesis, that the HSD group were selecting more positive reasons than the CONTROL group. Out of the 48 reasons that could be selected over the course of the 24 forecasts the HSD group selected a mean of 27.6 positive reasons and the CONTROL group a mean of 25.1 positive reasons (the difference was not significant). However, the number of positive reasons selected by the HSD group (only) was significantly greater than the 24 that would be expected if their selection had been random ($p< 0.0001$).

Discussion

Given the apparent lack of difference in the reported use of reasons between the two groups and also the similarity of their observed propensities to select positive reasons what might account for the more optimistic adjustments of the HSD group? One possibility is that for the HSD group, the positive reasons selected carried greater weight than negative reasons, while for the CONTROL group both positive and negative selected reasons carried equal weight. Indeed there may have been a discrepancy between the adjustment process (which may have been based largely on intuitive, unconscious, judgments) and the selection of reasons (which will have been deliberative and conscious). Thus the formal selection of reasons may not have been indicative of the way the reasons were being used in the forecast adjustment process. Valence priming would be more likely to be associated with intuitive reasoning.

Most surprisingly, however, while we have successfully replicated the apparent desirability bias observed in our company research, we have not replicated the optimism bias. Given that we had a mean promotion uplift of 80%, on average the forecasts of both groups would actually have been too pessimistic. What might account for this pessimism?

Recall that we provided a mix of reasons containing at least one negative reason. One possibility is that participants may have been trying to strike a balance between the effects of positive and negative factors so that the negative reasons were serving to depress the estimates. Indeed negative reasons may be more salient than positive which would be consistent with the negativity bias suggested by Krizan and Windschitl (2007). In the previously displayed promotion, participants will have no indication what the balance of positive and negative factors was and hence no idea that high uplifts might be achieved even when a negative factor is present. Our results may therefore simply show that desirability bias mitigates this negativity effect somewhat. These results may differ from our field observations because in companies no negative reasons may be sought or put forward.

There are other possible reasons for the lack of optimism. One is that the typical uplift of 80% is not salient - during the experiment the effect of only one past promotion is displayed on each graph.. However, for any given product, this is also likely to be the case in practice as promotions are likely to be relatively rare events. Another possibility is that because intervening sales observations are lower than the promoted sales: -they may have an anchoring effect leading to lower uplift estimates. However, this would also be the case in companies where optimism bias is observed despite this potential bias. A more plausible explanation is that our student subjects are unlikely to have had direct experience of promotions and they may perceive that 80% uplifts are unreasonably high, despite the brief that we gave them (though see Remus, 1986).

Our next experiment will try to address these issues. We will include 'all positive reasons' situations in addition to the 'all negative reasons' to see if a single negative reason, or a single positive reason, has undue influence on forecast adjustments. We will introduce a pre-experiment trial run to feedback to participants the factors that were associated with observed promotion uplifts so that they can see that greater 80% uplifts can be achieved despite the existence of one or more negative reasons. Also, we will include a report on one or more past promotions to link the reasons to the observed promotion effect. Finally, we need to test the existence of the bias when forecast accuracy is also rewarded and also test the use of reasons when they have manifestly different levels of reliability.

Conclusions

Desirability bias is an important factor in demand forecasting and can lead to significant forecast errors. It therefore has important implications for the design of forecasting support systems, in particular in relation to the way that information is displayed and used by forecasters. However, the link between information use and desirability bias is complex and requires further research. This will also examine the other possible mechanisms that are associated with desirability bias, such as differential scrutiny and confirmation bias and the ways that they might interact. Once the links between information use and desirability bias have been identified we will evaluate the aspects of support system design that may mitigate the bias, such as feedback (O'Connor et al, 2005), decomposition (Lawrence et al, 2006), guidance (Goodwin et al, 2011), database tools (Lee et al, 2007) and alternative ways of presenting information.

APPENDIX: Exit Questionnaire

1) Please rate your overall **level of knowledge** about demand forecasting

None				Excellent
1	2	3	4	5

2) Please rate your overall **level of confidence in your adjusted forecasts**

Not confident at all				Strongly confident
1	2	3	4	5

3) Please indicate your agreement/disagreement with the following statement:
"I have **carefully** examined the *time-series graphs*"

Strongly disagree				Strongly agree
1	2	3	4	5

4) Please rate **your** expected **forecasting performance** in this study

Very poor				Very good
1	2	3	4	5

5) Please indicate your agreement/disagreement with the following statement:
"I have **carefully** examined the *statistical forecasts*"

Strongly disagree				Strongly agree
1	2	3	4	5

6) Please rate your expectation for the **forecasting performance** of the **statistical forecasts**

Very poor				Very good
1	2	3	4	5

7) Please indicate your agreement/disagreement with the following statement:
"I believe that the **provided reasons are** very *intuitive*"

Strongly disagree				Strongly agree
1	2	3	4	5

8) Please indicate your agreement/disagreement with the following statement:
"I believe the **provided reasons** are very *clear to understand and use*"

Strongly disagree				Strongly agree
1	2	3	4	5

9) Please rate your assessment for the **value** of the **statistical forecasts**

Very poor				Very good
1	2	3	4	5

10) Please indicate your agreement/disagreement with the following statement:
"**Statistical forecasts** *had a direct influence* on **my own forecasts**"

Strongly disagree				Strongly agree
1	2	3	4	5

11) Please indicate your agreement/disagreement with the following statement:
"The **reasons** *had a direct influence* on **my own forecasts**"

Strongly disagree				Strongly agree
1	2	3	4	5

References

Fildes, R., Goodwin, P., Lawrence, M. and Nikolopoulos, K. (2009) Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning, *International Journal of Forecasting*, 25, 3-23.

Goodwin, P., (1996). Statistical correction of judgmental point forecasts and decisions, *Omega, International Journal of Management Science*, 24, 551-559.

Goodwin. P. Fildes, R., Lawrence, M. and Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega, International Journal of Management Science*, 39, 242-253.

Krizan Z & Windschitl, P.D. (2007) The influence of outcome desirability on optimism, *Psychological Bulletin*, 133, 95-121

Lawrence, M. Goodwin.P, O'Connor, M and Önkal, D (2006) Judgemental Forecasting: A review of progress over the last 25 Years, *International Journal of Forecasting*, 22, 493-518.

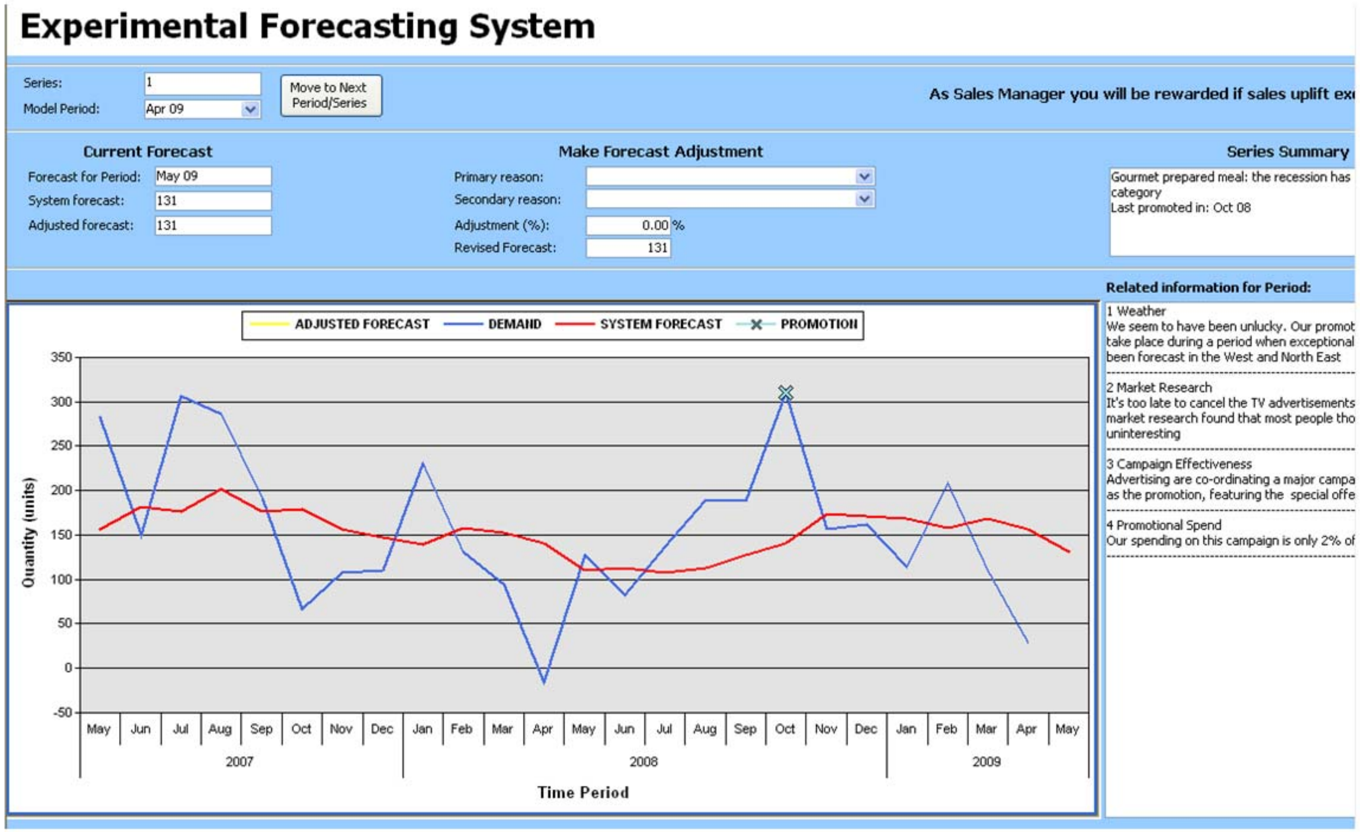
Lee, W.Y., Goodwin, P., Fildes, R., Nikolopoulos, K. and Lawrence, M. (2007) Providing support for the use of analogies in demand forecasting tasks, *International Journal of Forecasting*, 23, 377-390.

O'Connor, M, Remus, W. and Lim, K. (2005) Improving judgmental forecasts with judgmental bootstrapping and task feedback support, *Journal of Behavioral Decision Making*, 18, 247-260.

Önkal, D., Goodwin, P., Thomson, M., Gönül, M.S. and Pollock, A. (2009) The relative influence of advice from human experts and statistical methods on forecast adjustments, *Journal of Behavioral Decision Making*, 22, 390-409.

Remus, W.(1986) Graduate students as surrogates for managers in experiments on business decision making, *Journal of Business Research,* 14 (1986) 19-25.

Royzman, E.B. and Rozin, P. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5, 296-320**.**



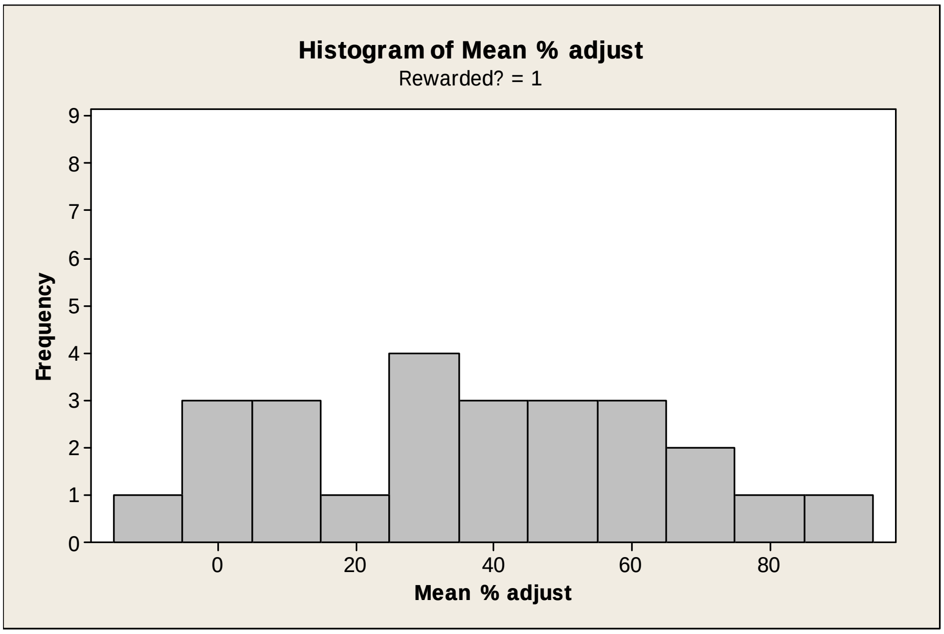
14

Figure 1 A typical screen display

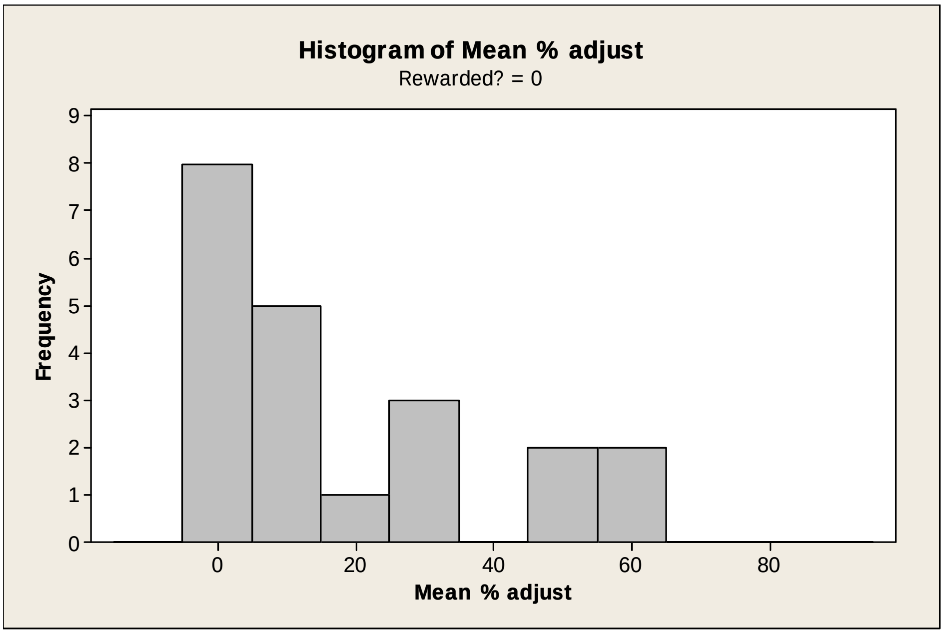
Reasons favouring >80% uplift	Reasons suggesting < 80% uplift
* These "buy-one-get-one-free" campaigns usually lead to a large uplift in sales	*These "3 for the price of 2" offers have not worked in the past for this type of product. People just don't want to buy in these quantities
*Focus groups discussing the proposed promotional pack thought the design excellent	*Focus group have been quite negative about the promotional packs
*Weather conditions in the Midlands where this product is popular, should help to boost sales substantially	*Weather factors are forecast to negatively impact on sales during the campaign in regions where this product usually sells well
*Supermarkets have agreed to display the product prominently during the campaign	* Unfortunately, supermarkets are not likely to display the product prominently during the campaign

13

Figure 2 Examples of positive and negative reasons



High Sales Desirable (HSD) Group



CONTROL group

Figure 3 Distributions of mean percentage adjustments made by participants