APPENDIX A

Method of Calculation for Rate of Weather Involvement in GA Fatalities

In the Introduction of this report, we stated "Data derived by us from National Transportation Safety Board statistics (1995-1997) support this assertion, showing IMC implicated in approximately 32% of GA fatalities." Here is how we calculated that figure:

Year	% GA accidents in VMC	% in IMC	% of IMC accidents fatal	% of VMC accidents fatal
1995	90	10	68	16
1996	90	10	63	15
1997	90	10	69	16
		average	66.7	15.7
		10%*66.7=	6.7	6.7 / 20.8 =
		90%*15.7=	14.1	0.32
		total % fatal	20.8	

We started with data from the National Transportation Safety Board (NTSB) *Annual review of aircraft accident data, U.S. general aviation*, calendar years 1995-97 (the latest available year was, indeed, 1997). The figure of 90% GA accidents occurring in VMC was taken from the text, which claimed "*More than* 90 percent of accidents occur in visual meteorological conditions" (emphasis ours, NTSB, 1997, p. 2 used to illustrate). So we assumed that 90% was an estimate, but a conservative one. We then followed these steps:

- The reports state that 90% of GA accidents occur in VMC (visual meteorological conditions).
- Therefore, by deduction, (100-90) = 10% must occur in IMC (instrument meteorological conditions).
- The reports state that 68, 63, and 69% of IMC accidents during their respective years involved fatalities (average = 66.7%), as opposed to 16, 15, and 16% of VMC accidents, respectively (average = 15.7%).
- If 10% of accidents involve IMC, and 66.7% of these are fatal, then (10% * 66.7%) = 6.7% of overall accidents therefore involve IMC PLUS fatalities
- If 90% of accidents involve VMC, and 15.7% of these are fatal, then (90% * 15.7%) = 14.1% of overall accidents therefore involve VMC PLUS fatalities.
- Therefore, ((6.7 / (6.7+14.1)) = 32%) is the ratio of (fatal accidents involving IMC / total fatal accidents), meaning that IMC is implicated in approximately 32% of GA fatalities.

APPENDIX B

Participant Debrief Form

S # ________

- ❑ What is your own normal personal minimum for VFR visibility? ________
- ❑ Your normal personal minimum for VFR cloud ceiling ________
- ❑ Are these minimums rock-solid, or do you adjust them a little, depending on the circumstances?________
- ❑ Have you ever flown this particular route before (or a similar situation)? ________
- ❑ Did the distance you had to fly through bad weather affect your willingness to take off ? ________ (for example, if the distance had been greater, would you have been even less inclined to take off than you were?)
- ❑ If you were in the "high-incentive" condition, did this affect your willingness to take off? ________
- ❑ Do you think having passengers would affect your willingness to take off? (increase it ____, no change____, decrease it ____)
- ❑ If you had a lot more flight hours, would that have change your willingness to take off? (increase it ____, no change____, decrease it ____)
- ❑ If your flight mission had been critical (for example, delivering a human heart for surgery), would that change your willingness to take off? (increase it ____, no change____, decrease it ____)
- ❑ Have you ever flown a Piper Malibu before?_____ Did this affect your willingness to take off?
- ❑ It made me **more** willing because I was anxious to try it out ___,
- ❑ It didn't matter one way or the other ___,
- ❑ It made me **less** willing because I was afraid I'd make more mistakes ___
- ❑ Did the fact that this was a simulation (and not reality) affect your willingness to take off?
- ❑ It **increased** willingness because
  - ❑ (a) I wanted to fly the sim___ and/or
  - ❑ (b) I knew I couldn't really get injured in it___,
- ❑ No, it had no effect because
  - ❑ (a) it didn't matter to me one way or the other___
  - ❑ (b) there were positives and negatives but they cancelled each other out___
- ❑ It **decreased** willingness because
  - ❑ (a) I was unfamiliar with this particular simulator___
  - ❑ (b) I didn't want to make any mistakes in front of the experimenter___
- ❑ How economically significant was the money to you?
  1__not at all 2__a little 3__fairly significant 4__significant 5__very significant
- ❑ If you were to crash in the simulator, how embarrassed would you be?
  1__not at all 2__a little 3__fairly 4__significantly 5__extremely
- ❑ Have you ever had a bad flight experience related to weather?___If so, please describe briefly below.

APPENDIX C

	YrsFlg	age	gender	Type Lic	Inst rating	fhtot	fh12mtm	fh90dtm	HP Tot	HP 12m Tm	HP 90d Tm	SI time	SI 12m Tm	SI 90d Tm	AI Time	AI 12m Tm	AI 90d tm
N	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
# Missing																	
Median	1.8	23.5	male	priv.	50/50	183.5	80.0	35.0	5.0	3.0	0.0	25.0	10.0	5.0	2.0	1.0	0.0
Average	4.3	26.1				753.0	159.9	49.1	37.9	16.9	6.9	69.3	20.7	14.4	52.7	4.7	2.2
S.D.	7.6	8.5				2605	239.6	54.1	88.5	40.8	18.8	256.8	30.1	26.5	322.3	11.0	5.2
Max	48.25	69				20000	1200	250	560	200	125	2000	175	160	2500	75	25
Min	0.25	18				35	0	0	0	0	0	0	0	0	0	0	0
Skew	4.2	2.8				7.1	2.9	1.9	4.2	3.6	4.9	7.4	3.4	3.8	7.7	4.9	3.1
SE Skew	0.31	0.31				0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
z_{skew}	13.67	9.07				22.95	9.47	6.05	13.51	11.57	15.84	24.13	11.15	12.29	24.86	15.90	10.09
p_{zskew}	0.000	0.000				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kurtosis	21.1	10.6				52.8	9.0	3.5	21.0	12.4	27.4	56.8	14.5	17.3	59.2	29.1	9.5
SE Kurt	0.61	0.61				0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61
z_{kurt}	34.8	17.5				86.8	14.8	5.7	34.5	20.4	45.1	93.3	23.9	28.4	97.4	47.8	15.5
p_{zkurt}	0.000	0.000				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Outlier(s)?	yes	yes				yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

	Vismin	ceilmin	minadj	flnrt	dthruwx	buckmot	passmot	flthrmot	missnmot	malibu	malibsub	simmott	simmosub	bucksig	crashsig	trafsig	badwx
N	59	59	57	57	57	28	56	55	47	53	50	51	46	51	40	11	55
# Missing	1	1	3	3	3	32	4	5	13	7	10	9	14	9	20	49	5
Median	4.0	2000	1	0	1	1	-1	0	1	0	0	0	3	3	3	2	0
Average	4.1	2186	0.8	0.4	0.6	0.6	-0.5	0.3	0.7		-0.2	0.2		2.7	3.2	2.3	
S.D.	1.6	1213					0.7	0.6	0.5		0.6	0.5		1.3	1.2	1.3	
Max	10	8000	1	1	1	1	1	1	1	1	1	1	6	5	5	5	1
Min	1	1000	0	0	0	0	-1	-1	-1	0	-1	-1	1	1	1	1	0
Skew	1.2	2.4					1.0	0.0	-1.3		0.0	0.4		0.2	0.2	0.8	
SE Skew	0.31	0.31					0.32	0.32	0.35		0.34	0.33		0.33	0.37	0.66	
z_{skew}	3.75	7.77					3.08	0.02	-3.73		0.00	1.34		0.53	0.43	1.22	
p_{zskew}	0.000	0.000					0.004	0.399	0.000		0.399	0.163		0.347	0.363	0.190	
Kurtosis	2.5	8.8					-0.1	-0.4	0.7		-0.1	0.4		-1.1	-1.2	0.5	
SE Kurt	0.61	0.61					0.63	0.63	0.68		0.66	0.66		0.66	0.73	1.28	
z_{kurt}	4.0	14.4					-0.2	-0.6	1.0		-0.2	0.6		-1.7	-1.6	0.4	
p_{zkurt}	0.000	0.000					0.390	0.324	0.235		0.389	0.339		0.092	0.108	0.367	
Outlier(s)?	yes	yes							yes								

Standard z-tests (e.g. skew/[standard error of skew]) showed that the demographic data were greatly skewed by the presence of a small number of older pilots with, for instance, a great deal of flight experience. Winsorizing corrected virtually all this non-normality. Appendix D explains the factors examined.

APPENDIX D

Complete list of factors examined

Name	Description	Name	Description
subjnum	Order in which S was run	asi	AIS Anxiety Sensitivity Index
idnum	S. ID number	bis	BIS-10 Barratt Impusiveness Scale total
takeoff	Takeoff (yes/no)	imp_plan	BIS Impulsive Planning
latcy	Latency (time elapsed before takeoff)	imp_motr	BIS Motor Impulsivity
vis	Ground Visibility (statute miles)	imp_cog	BIS Cognitive Impulsivity
ceil_k	Ceiling (in thousands of ft)	impuls	EIS Eysenck Impulsivity Scale Impulsivity
incent	Financial Incentive (bonus / no bonus)	ventur	EIS Venturesomeness
wxsevrty	Weather severity (1 / (Vis x Ceil))	empath	EIS Empathy
exptr	Experimenter (WK or HH)	hei	HEI Hazardous Events Index (Hunter)
yr_flg	Year started flying		Multidimensional Personality Questionnaire
yrs_flg	Years flying	wellbe	MPQ Wellbeing
age	Age	socpot	MPQ Social Potency
gender	Gender	achieve	MPQ Achievement
type_lic	License type	socclose	MPQ Social Closeness
inst_rtg	Instrument rating (yes/no)	stress	MPQ Stress Reaction
fh_tot	Flight hours total	alienate	MPQ Alienation
fh_12m	FH past 12 months	aggress	MPQ Aggression
fh_90d	FH past 90 days	control	MPQ Control
hp_tot	High-performance aircraft hours total	harmav	MPQ Harm Avoidance
hp_12m	HP past 12 mo	tradit	MPQ Traditionalism
hp_90d	HP past 90 days	absorpt	MPQ Absorption
si_tot	Simulated instrument hours total	roq_c	ROQ Risk Orientation Q'naire Cautiousness
si_12m	SI past 12 mo	roq_p	ROQ Risk Propensity
si_90d	SI past 90 d	sss	SSS Sensation Seeking Scale
ai_tot	Actual instrument hours total	anx_st	STAS State-Trait Anxiety Scale State
ai_12m	AI past 12 mo	anx_tr	STAS Trait
ai_90d	AI past 90 d		Balloon Analog Risk Task (Lejuez)
vis_min	Personal visibility minimum	durn_sec	BART Task Duration (seconds)
ceil_min	Personal ceiling minimum	pumpsavg	BART Average # of pumps
min_adj	Adjust mins. to match the situation?	pmpavglo	BART Average # of pumps (low incentive)
fln_rt	Flown this route before?	pmpavgme	BART Average # of pumps (med incentive)
dthruwx	Did distance through the weather matter?	pmpavghi	BART Average # of pumps (hi incentive)
buck_mot	Was the $ bonus a motivation? (yes/no)	pmpadjav	BART Adjusted Ave. # of pumps
pass_mot	Would passengers have been a motivation?	padjavlo	BART Adj. Ave (low incentive)
fhincmot	Would more flt hrs increase motivation?	padjavme	BART Adj. Ave (med incentive)
missnmot	Was the type of mission a motivation?	padjavhi	BART Adj. Ave (high incentive)
mal_sub	Was the type of flight simulator a motivation?	pay_tot	BART Total Payoff (cents)
sim_mot	Was fact of being a sim (vs. reality) a motvn?	pay_low	BART Total Payoff (low incentive)
simmotsb	Sub-categories of sim_mot	pay_med	BART Total Payoff (med incentive)
buck_sig	How significant was the $$ to you?	pay_hi	BART Total Payoff (high incentive)
crashsig	Was worrying abt crash a motivation?	bang_tot	BART Total Balloon Explosions
tx_mot	Was traffic a motivation?	bang_low	BART Balloon Explosions (low incent)
badwx	Ever had a bad wx experience? (y/n)	bang_med	BART Balloon Explosions (med incent)
asa	Aviation Safety Attitude Scale (Hunter)	bang_hi	BART Balloon Explosions (high incent)

Three of these factors were not predictors, namely *idnum*, *takeoff*, and *latcy*. *IDnum* was merely the numerical proxy for subject name. *Takeoff* was the dependent variable. *Latency* was a descriptor, and could not be used as a discriminative predictor because maximum latency (120 minutes) was always associated with takeoff, anything less, with non-takeoff. Therefore, there were 83 usable predictors.

APPENDIX E

Statistical Issues in Logistic Regression

Outliers. *Outliers* are defined for our purposes here as any score greater than 3 standard deviations above or below the mean. Outliers can sometimes exert an almost unbelievable effect on the statistical outcome of an analysis. Take, for example, a distribution of ones and zeros representing Financial Incentive, one of our predictors of Takeoff. For the full data set, N=60, our actual raw distribution yields the following result during SPSS logistic regression:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	INCENT	.981	.541	3.287	1	.070	2.667
	Constant	-.847	.398	4.523	1	.033	.429

a. Variable(s) entered on step 1: INCENT.

This result says that the probability of Incentive being a significant predictor of Takeoff is .070.

Now let us change *one single value* in the data distribution from a "0" to a "10" to represent, say, a typographical error during data coding. Changing just this *one value in 60* results in the following:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	INCENT	-.027	.204	.018	1	.894	.973
	Constant	-.319	.294	1.171	1	.279	.727

a. Variable(s) entered on step 1: INCENT.

Suddenly we have gone from $p = .070$ to $p = .894$ in one step—by turning a single data point into a gross outlier. Obviously, this says a lot about the need for accurate data coding. It also says quite a bit about how outliers can affect an otherwise normal data distribution. Now logistic regression does not have an underlying logical assumption of normality (Tabachnick & Fidell, 2000). You could, for instance, use data with any relatively symmetrical distribution. But it does have problems with outliers, as this clearly demonstrates.

The data in this study showed outliers in the demographics, where a small number of older pilots significantly skewed the distributions for predictors such as age, flight hours, and years flying. Without some kind of correction, therefore, the effect of outliers would have led us to seriously misinterpret the statistical analysis.

Applying a data transformation (such as a square root or logarithmic function) is a common way to deal with outliers. A somewhat less well-known, but equally respected treatment is *winsorization* (Winer, 1971, pp 51-54). 1971). In winsorization, the two most-extreme values in the distribution (the one highest and the one lowest) are replaced by a copy of the next most-extreme values. For example, in the distribution

0 1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5 5 5 99 (mean 7.23, SD 20.54)

we would replace the "0" with a "1" and the "99" with a "5."

1 1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5 5 5 5 (mean 3.00, SD 1.38)

Now this new distribution is still not normal because it is too flat. But it no longer has the gross outlier it once had. That extreme value of "99" is still represented by a relatively high value, which preserves the ordinality (rank order) of the scores. But notice that there was no actual change to most of the numbers. Only two values were changed, and one of those was a very modest change from a "0" to a "1." Whereas, if we had applied a mathematical function such as a square root to shrink the "99" closer to the mean, almost all of the values would have been affected. Here winsorization exerts its biggest effect on the greatest offender, which is exactly how data conditioning should work. This illustrates how this technique can sometimes preserve the spirit and actuality of a distribution much better than can some of the more routinely used methods. For this reason, it was the method of choice for our data.

If a distribution has more than one outlier, say

0 0 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5 5 47 99 (mean 9.09, SD 22.23)

we simply apply the winsorization procedure twice, to yield

1 1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5 5 47 47 (mean 6.82, SD 13.06)

at stage one and

2 2 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5 5 5 5 (mean 3.18, SD 1.14)

at stage two. In this example, the two-stage winsorization affects 6 values, rather than just 2. For this reason we have to be careful in repeating this process too often, since it can lead to the antithetical problem of range restriction.

In this, study winsorization was limited to no more than 2 stages. For example, in the full data set (N=60) 16 demographic variables were seen to have outliers > 3 SD, and therefore received either a 1- or 2-stage winsorization, depending on what was needed to eliminate these outliers. After treatment, all 16 variables emerged corrected to tolerance.

A final point worth mentioning is that winsorization has a net result of making our statistical analysis more conservative. This happens precisely because the distributions' ranges and variances contract during conditioning, and any time variance contracts, *p*-values generally contract as well. This is not true with purely ordinal statistics, because these calculate their value based on nothing more than rank order. But both chi-square and logistic regression do not fall into that category. While logistic regression is often touted as being distribution-free, in fact, we have graphically illustrated that things are a bit more complex. Outliers skew its innermost calculation of likelihood ratios (SPSS, 2004). However, the data conditioning process employed here allowed us to successfully treat data and to present *p*-values representing useful-yet-conservative estimates of statistical reliability.

Correction for Familywise Error. Another important issue is the one of correcting *p*-values to account for the number of predictors examined. Most statisticians recommend some sort of correction for experimentwise Type I error (unwarranted rejection of the null hypothesis). Otherwise, if we do many tests, odds are that some will be "significant" simply by chance.

However, we consciously chose to deviate from that standard procedure because, in an exploratory study such as this, such rigor, while admirable in one sense, would most certainly have the net result of too much Type II error, that is, failure to detect a true effect where there was one. And, while the danger of inflated experimentwise Type I error was fully appreciated, we also felt it made more sense to report low *p*-values where found, because these really do represent the best guess we have regarding effect.

The ideal way to resolve the problem, of course, is to run Monte Carlo simulations to get estimates for mean predictivities and R^2s, given specific parameters of specific models. This was done in Part II of this report. Another accepted approach is to replicate studies or parts of studies, using different participants. That will be done in follow-up studies, whenever possible.

APPENDIX F

Brief Description of Logistic Regression

Logistic regression is a statistical technique specially constructed for use with discrete dependent variables, for example, Takeoff versus No Takeoff. It is a very useful technique, but it is also extremely easy to miscode, misunderstand, and misinterpret. The best way to understand it is through a combination of mathematics and example.

Regression is the search for factors that predict other factors. In this experiment, we wanted to predict the likelihood that an average pilot would take off into known marginal weather, given the added influence of financial incentive. Three of our predictive factors (*Visibility, Cloud Ceiling,* and *Financial Incentive*) were under experimental control; the rest reflected either demographic or personality characteristics of each individual pilot.

Logistic regression uses an equation to predict the outcome of an *event*, in this case Takeoff versus No Takeoff (Dreyszig, 1972; Norušis, 1999; SPSS, 2004). This equation is

$$p_{event} = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_n X_n)}} \qquad (1)$$

where e is the natural log (approximately 2.718), B_0 (beta-sub-zero) represents a constant, and B_i is the corresponding beta weight for the ith predictor, X_i score. Varying the values of the exponent of e produces a distinctive sigmoid (S-shaped) curve capable of representing probability of takeoff

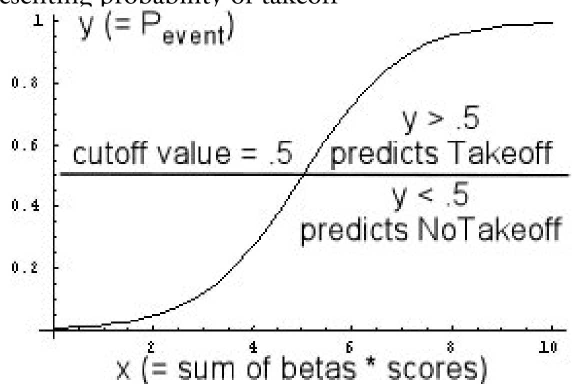


Figure F1. Sample plot of the sigmoid function $y=1/(1+ e^{-(-5 + x)})$, showing how the overall value of the prediction equation lies between zero and one. In our case, this represents the probability of a pilot taking off, 0-1 (0-100% chance), given some particular combination of predictor scores X_1 through X_n. When a given pilot's calculated probability exceeds an predetermined cutoff level (for example, 0.5), we will predict "Takeoff," otherwise we will predict "No Takeoff."

Logistic regression has two very attractive advantages over competing statistics. First, as we mentioned, it allows us to make predictions. Second, it allows us to test statistical interactions between predictors. Equation 2 shows how this is typically implemented, showing the prediction equation with its constant B_0, one main variable X_1, plus one interaction term involving three factors B_2, X_2, and X_3. Notice that the interaction term literally involves multiplying together the separate predictors. This is an important point to which we will presently return.

$$p_{event} = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2 X_3)}} \qquad (2)$$

To illustrate this numerically, suppose we tested a model where X_1 represented a pilot's score of 34 on the Rohrmann Risk Orientation Questionnaire (subscale P), and where X_2 and X_3 represented the interaction of Visibility x Ceiling, 3 (miles) and 1 (feet, in thousands), respectively. In that case, the prediction equation for that individual would be

$$p_{Takeoff} = \frac{1}{1+2.718^{-(-3.396+.074(34)+.197(3)(1))}} = .428$$

Since .428 is less than the default cutoff value of .500, we would predict that this particular pilot would not take off.

When we run the SPSS analysis on the full data set, the program basically goes through a similar process for each individual, computing a set of guesses regarding each pilot's takeoff. Some guesses will be right, others wrong. Then the beta weights are shifted slightly, the analysis is repeated, and the results compared to the priors. If shifting the betas in that direction produces improvement, the direction of shift is repeated, otherwise it is reversed. After a certain number of iterations, the process halts and summary tables are produced. Here is one of the summary tables for an actual model:

The most important numbers in this table, as far as we are concerned, are the β weights, and the significance of the Wald statistic (Sig). What the βs here tell us is primarily the direction of the association between a predictor and the outcome. Take ROQ_P, whose β is positive. That tells us that an *increase* in the ROQ_P score predicts an *increase* in takeoff probability. If β had been negative, an increase in the ROQ_P score would have predicted a *decrease* in takeoff probability. The magnitude of β is also important, though its interpretation is more complicated. We have to take into account how the predictor was scored. Since β is multiplied by the predictor score, if the predictor scores are large (e.g. 1000 feet altitude) then even a small β can be very influential.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	ROQ_P	.074	.061	1.465	1	.226	1.077
	CEIL_K by VIS	.197	.095	4.325	1	.038	1.217
	Constant	-3.396	1.902	3.186	1	.074	.034

a. Variable(s) entered on step 1: ROQ_P, CEIL_K * VIS .

The Wald significance (Wald p value) works very much like a normal statistical p value. Wald p tells us the reliability of the measurement, estimating the proportion of times we would expect to find a different result, if we repeated the analysis a large number of times. In this particular instance, ROQ_P's Wald p is .226—too large to be considered reliable.

The Constant (β_0) in this analysis behaves somewhat like other predictors. However, the Constant is sometimes the most difficult term to interpret in a regression model. It can reflect the sample's base rate for the dependent variable. However, this depends on what other predictors happen to be in the model. If all the other predictors are "Go" predictors (ones with $\beta > 0$, where an increase in predictor score reflects an increase in the DV), then the Constant may take on a contrarian role and assume $\beta < 0$). Whereas, had all the predictors been "No-go," with $\beta < 0$, then the Constant may have a $\beta > 0$. In mixed models, with both Go and No-go predictors, things could go either way. Therefore, interpretation of the Constant has to be approached with skill and caution.

Categorical Variables and the Use of Contrasts

The analytic usefulness of logistic regression is a big plus. What is not a plus is the meticulous care that has to go into coding the data, setting up the analysis, and interpreting the results.

For one thing, the technique is susceptible to outliers, as we mentioned. Misentry of even a single data point can wreck an analysis.

Another serious difficulty lies in the use of categorical predictors. Although logistic regression is technically capable of handling both categorical and continuous variables, special care needs to be taken when using categoricals. As long as all variables are continuous, either ordinal or ratio-scale, no special care needs to be taken. But categoricals are different. This is because the program takes categoricals coded as letters and converts them internally into zeros and ones. For example, we had two experimenters involved in running the participants. Call them "H" and "B." During the SPSS analysis, experimenter "H" is internally converted by the program into either a zero or a one, in order to be plugged into equation 2. This conversion introduces the opportunity for serious conceptual errors to be made if we are not scrupulous in coding in the data, thinking out our analysis, and interpreting the results.

To drive this idea home, let us take this example further. If, during the analysis, we fail to specify the variable EXPTR as categorical (which requires bringing up a dialog box and making some adjustments), then we could be making a large mistake. That is because SPSS has automatic defaults and will change *any* letter into a number, whether or not we understand what it is doing. So look at the equation—trying to treat "H" as "nothing" and "B" as "one unit of something" makes sense only in a very limited context. And, say we run a model containing an interaction. What the mathematics actually does is eliminate the effect of ALL the predictor scores in that interaction term whenever it calculates a data point involving "H," because it multiplies the other variables in that interaction term by zero for that data point. And this is something we might not have intended to do exactly that way. This is the way we do contrasts, but the point is that the program can be doing a contrast we do not know it is doing if we do not understand exactly what is happening mathematically.

Looking at some actual SPSS output will make this a little clearer. Below is some output for the simple model EXPTR + Constant. In the first case, "H" was set to internally code as "1" and "B" as "0." In the second case, those codes were reversed.

You can see that the statistical significance (Sig.) of EXPTR does not change, and that this particular variable did not produce a reliable effect (.593). The betas for EXPTR are the same, just with opposite sign. This is simply because we are logically testing one thing "A" against another thing "not A" and, because there are only two things, so "not A" has no degrees of freedom. But, as we would expect from the math and the iterative computational algorithm we talked about, the constants turn out to be different. This is because, in the first instance, "H" was exerting the

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	EXPTR(1)	-.288	.539	.285	1	.593	.750
	Constant	-.223	.335	.443	1	.506	.800

a. Variable(s) entered on step 1: EXPTR.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	EXPTR(1)	.288	.539	.285	1	.593	1.333
	Constant	-.511	.422	1.468	1	.226	.600

a. Variable(s) entered on step 1: EXPTR.

primary mathematical effect, since B = 0, whereas B = 1 in the second. But the constant is being updated always. Consequently, the two models are related, and similar, but the first one is really measuring "the effect on takeoffs of a pilot's being run by 'H'," as opposed to being run by 'not H' (i.e. 'B')." The negative beta in the first case means "pilots were less likely, on the whole, to take off if they were run by 'H' than by 'not H'" (although recall that p is not reliable, so we would not ultimately assert any difference) In this case, it so happens that 'not H' has to mean "B," but that was only because there were only two experimenters. Had there been three, we would have had to test a third contrast, and each would have tested primarily the effect of that one experimenter, set up consciously by us to code as "1."

Things get even more interesting when it comes to interactions involving more than one categorical variable. The essential logic remains the same, however: a) contrasts focus on whatever happens to be coded "1," and b) interactions go to zero whenever any single term in them becomes zero. The bottom line is that we cannot simply mindlessly run SPSS and hope to understand the data.

Problems Associated With Logistic Regression

Like all statistics, logistic regression is not a perfect technique (Tabachnick & Fidell, 2000). Some of its weaknesses include

1. *Correlation does not imply causation.* All regression techniques do is to establish a mathematical relation between the presence/absence of one thing and the presence/absence of another. But such correlation does not necessarily mean, for instance, that Factor A *causes* Factor B. The classic counterexample is the case where Factor A and Factor B are both caused by Factor C. In that case, A and B still show correlation, but there is no causation whatsoever between A and B.

2. *Outliers can greatly skew models and parameter estimates.* We demonstrated this clearly in Appendix E. Fortunately, this problem was easily overcome by winsorizing the data.
3. *Independence of samples is assumed.* Logistic regression is basically a between-subjects technique, not for repeated measures gathered over time. That was not a factor in this study, however.
4. *Absence of multicollinearity is assumed.* If predictors are highly correlated, they are probably measuring the same factor, and will not contribute much, if anything additional to a model, other than wrongly inflated significance. Fortunately, the models we present did not pose this problem (see Appendix G for the intercorrelation matrices).
5. *The ratio of cases to model predictors is important.* A common rule of thumb, seen in many textbooks, is that a model should contain no more than one predictor per 10 cases (e.g., per 10 pilots). If a constant is used, this should be counted as one predictor However, we noticed an ancillary problem during this analysis, namely
6. *The case-to-predictor ratio issue extends to the number of predictors measured <u>before</u> analysis is commenced.* This is discussed in greater depth below, and in the Part II report.

Problems Associated With Too Many Predictors in Forward Stepwise Logistic Regression

At some point, we had the intuition that simply trying to examine too many predictors in our primary technique of forward stepwise regression could introduce a combinatorial problem. That theoretical problem is easiest illustrated using our actual situation. We started with 83 candidate predictors, some of which were eventually eliminated due to reasons such as having missing values or being discrete (which often led to unwieldy combinations of contrasts). So, in the end, we looked at roughly 60 predictors.

Now, consider the following deductive logic: Suppose you were trying to model some data taken from 30 pilots, upon whom you had 60 measurements (predictors) each. This would correspond to, say, our Low Financial Incentive group. Then the rule of thumb we mentioned above in Point 5 suggests that all such models should have no more than 30/10 = 3 predictors. So far, so good.

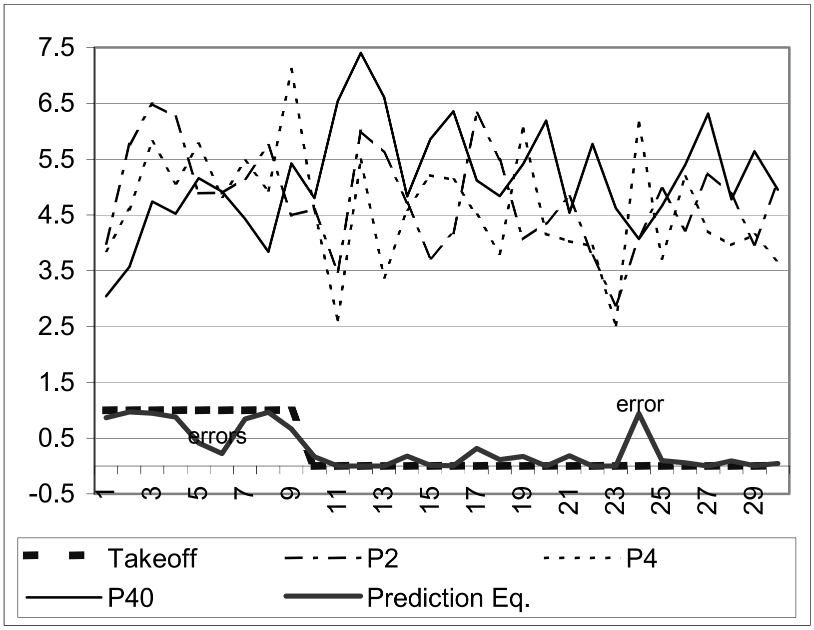
The problem comes when we consider random numbers. Suppose every one of our predictors was simply "noise," taken randomly from a Gaussian (normal, bell-shaped) distribution of numbers. Given that the logistic regression prediction equation is basically

$$P_{event} = \frac{1}{1+e^{-((\beta_0+\beta_1 X_1+\beta_2 X_2(+\beta_3 X_3)))}} \tag{1}$$

notice how the exponent term $-((\beta_0 +) \beta_1 X_1 ...)$ is really a sum. It will be the sum of our predictors (each weighted). That means that, whatever the actual numbers are for each pilot's predictor scores, we are going to weight them, then add them up to form a total, which will then be plugged into Eq. 1. So what are the chances that, <u>given nothing but random numbers</u>, SPSS will ultimately end up finding the precise set of β weights such that the Equation 1 turns out greater than 0.5 for pilots who subsequently took off, versus a predicted score of less than 0.5 for those who did not?

Shockingly, the answer is that it is highly likely. We verified this by running Monte Carlo simulations, a standard technique in statistics. Using normal random number generation with μ (mu, mean) of 5 and σ (sigma, standard deviation) of 1, we were easily able to duplicate results such as the following:

This illustrates that SPSS essentially "made sense out of nonsense." It summed the three random pseudo-predictor scores for each pilot, shown by the three jagged curves, multiplying each score by the β weights it derived, inserted them into Equation 1 and came up with the much-more regular solid "Prediction Equation" line. Notice how closely that matched the thick, dashed "Takeoff" line representing a dependent variable score of 1 for a takeoff and 0 for a non-takeoff. The three points where those two curves did not closely correspond are labeled as "error." Since 27 of the 30 cases were "predicted" correctly, this model's predictivity was .90.



Why is this? Well, look at it from the model's point of view. In forward stepwise regression, the software first chooses one predictor to enter into the model, then a second, a third, and so forth. With 60 predictors, it has 60 candidates for the first choice, 59 for the second, 58 for the third, and so on. Thus, in a three-single predictor model, there are 60*59*58/(3*2*1) = 34,220 possible combinations,[1] not even counting interactions. What is happening is that, given such a huge number of combinations, at least one is highly likely to form a highly "predictive" sum, even though, purely taken individually, none of the individual scores has any meaning whatsoever.

We did 100 Monte Carlo simulations for each of our low- and high-incentive groups, with and without a constant in the model. While this was well under the usual standard of 1000-10000 or so simulations per condition, doing each simulation was quite tedious, and these 400 runs did have sufficient reliability to illustrate our basic points.

Here we see that the proportion of takeoffs matters. Noise models with a proportion of takeoffs close to .5 show lower predictivity and Nagelkerke R^2 than ones with a proportion of takeoff equal to .3. But, overall, predictivities were still in the 70-90% range, and R^2s in the 40-70% range for these random-number models.

Evaluation of the Meaningfulness of Our Data

So how reliable were the conclusions for this Part I report?

	Low Fin. Incentive group models with constant		High Fin. Incentive group models with constant	
	Predictivity	Nagel R^2	Predictivity	Nagel R^2
$\mu_{MonteCarlo}$	80.4	0.36	76.3	0.48
CI 95	≅.89	≅.59	≅.85	≅.64
$\mu_{ActualData}$	85.7	0.52	75	0.28
α estimated	0.16	0.08	NS	NS

The method used to derive these estimates is detailed in the companion report *Pilot willingness to take off into marginal weather, Part II: Antecedent overfitting with forward stepwise logistic regression.*

[1] The reason for the denominator is that the order of terms in the model makes no difference. SPSS logistic treats "ABC" the same as "ACB," "BAC," "BCA," "CAB," and "CBA"—three degrees of freedom for the first choice, two for the second, and one for the last.

To summarize, the .95 confidence intervals around the predicted means (CI .95) imply that any model exceeding these estimates for predictivity and R^2 is highly likely to be a better-than-chance model. Confidence intervals are a standard approach used in many statistics.

The full data set and high incentive models derived from real pilots' data in the current study did not differ significantly from what could be expected from random number simulations. That is why we limited our primary observations to high-level conclusions and the Low Incentive data. The real-pilots' low-incentive 85.7% predictivity did exceed the random-generated Monte Carlo mean of 80.4%, although it did not top the estimate of 89% for the .95 CI. Their Nagelkerke R^2 of .52 considerably bested the Monte Carlo mean of .36 and came close to meeting the .95 CI of .59. So, judging from the Monte Carlo scatterplots (shown in Part II), reliability for the low incentive n=28 experimental data was roughly α =.16 for predictivity and α =.08 for R^2.

As said previously, for the purposes of a preliminary report such as this, it is often wiser to be somewhat relaxed in reporting results than we would be later on in the research process. This is because of the Type I-Type II error tradeoff, that is, where excessive stringency in setting significance levels results in a lower number of false positive results but strictly at the cost of a higher number of missed results. In other words, at first the strategy involves going for breadth of findings. The small number of results that fail to be reliable will be discovered and eliminated as other studies cross check results presented here.

APPENDIX G

	Years flying	Flight hrs tot	FH past 12 mo	FH past 90 d	Hi perf hrs tot	HP past 12 mo	Simd instr hr tot	SI past 12 mo	Act inst hr tot	AI past 12 mo	Anxiety Sensitivity Index	BIS Impulsive Planning	BIS Motor Impulsivity	BIS Cognitive Impulsivity	BIS Total	EIS Impulsivity
Age	0.686															
FH past 12 mo		0.619														
FH past 90 d			0.737													
Hi perf hrs tot		0.591	0.482													
HP past 12 mo			0.631		0.744											
HP past 90 d					0.541	0.767										
Simd instr hr tot		0.673	0.686	0.612												
SI past 12 mo				0.520			0.567									
SI past 90 d				0.517				0.828								
Act inst hr tot		0.904	0.481		0.581		0.642									
AI past 12 mo			0.693	0.656			0.559		0.500							
AI past 90 d			0.623	0.605						0.915						
BIS Cognitive Impulsivity												0.560	0.601			
BIS (tot)												0.815		0.775		
EIS Impulsivity												0.558	0.578		0.488	
Hazardous Events Index		0.604			0.508				0.597							
MPQ Control												-0.576	-0.576		-0.488	-0.766
MPQ Harm Avoidance																
Rohrmann Cautiousness																
Rohrmann Risk Propensity																
Sensation Seeking Scale												0.487	0.532			
STAI State Anxiety											0.532					
STAI Trait Anxiety											0.587			0.578		
BART pumps (ave lo pay)																
BART pumps (ave med pay)																
BART pumps (ave hi pay)																
BART pumps adj ave																
BART pumps adj ave(lo)																
BART pumps adj ave (med)																
BART pumps adj ave (hi)																
BART ave pay (cents)																
BART ave pay (lo)																
BART ave pay (med)																
BART ave pay (hi)																
BART explosions (tot)																
BART explosions (lo)																
BART explosions (med)																
BART explosions (hi)																

	EIS Venturesomeness	MPQ Stress Reaction	MPQ Control	MPQ Harm Avoidance	Rohrmann Risk Propensity	STAI State Anxiety	BART pumps (ave)	BART pumps (ave lo pay)	BART pumps (ave med pay)	BART pumps (ave hi pay)	BART pumps adj ave	BART pumps adj ave (lo)	BART pumps adj ave (med)	BART pumps adj ave (hi)	BART ave pay (cents)	BART explosions (tot)	BART explosions (lo)	Wald p (model w. constant)
Age																		
FH past 12 mo																		
FH past 90 d																		
Hi perf hrs tot																		
HP past 12 mo																		
HP past 90 d																		
Simd instr hr tot																		
SI past 12 mo																		
SI past 90 d																		
Act inst hr tot																		
AI past 12 mo																		
AI past 90 d																		
BIS Cognitive Impulsivity																		
BIS (tot)																		
EIS Impulsivity																		
Hazardous Events Index																		
MPQ Control																		
MPQ Harm Avoidance	-0.567																	
Rohrmann Cautiousness			0.532															
Rohrmann Risk Propensity	0.562																	
Sensation Seeking Scale	0.647		-0.520	-0.484	0.493													
STAI State Anxiety																		
STAI Trait Anxiety		0.648				0.669												
BART pumps (ave lo pay)							0.845											
BART pumps (ave med pay)							0.937	0.725										
BART pumps (ave hi pay)							0.829		0.732									
BART pumps adj ave							0.982	0.810	0.924	0.831								
BART pumps adj ave(lo)							0.822	0.967	0.707		0.818							
BART pumps adj ave (med)							0.908	0.679	0.964	0.740	0.930	0.673						
BART pumps adj ave (hi)							0.793		0.701	0.977	0.813		0.726					
BART ave pay (cents)							0.554		0.523	0.574	0.576			0.538				
BART ave pay (lo)							0.528	0.581	0.496		0.580	0.712			0.672			
BART ave pay (med)															0.805			
BART ave pay (hi)														0.512	0.697			
BART explosions (tot)							0.913	0.837	0.858	0.684	0.879	0.777	0.844	0.655				
BART explosions (lo)							0.749	0.878	0.623		0.677	0.775	0.568			0.841		
BART explosions (med)							0.761	0.729	0.789		0.744	0.693	0.755			0.890	0.681	
BART explosions (hi)							0.753		0.705	0.832	0.759		0.769	0.780		0.735		

These reflect the relatively young age of the pilots

Intra-BIS correlations

Intra-BART correlations

Intercorrelation matrix, Pearson R, all Ss (N=60), all variables with significance $p < .0001$ ($R^2 > .23$).

0.01 Threshold significance **(p<θ)** 2-tailed

	Years flying	Age	Flight hrs total	FH past 12 months	FH past 90 days	Hi perf hrs tot	HP past 12 mo	HP past 90 d	Simd instr hr tot	SI past 12 mo	SI past 90 d	Act inst hr tot	AI past 12 mo	ASA	ASI	BIS Impulsive Planning	BIS Motor Impulsivity	BIS Cognitive Impulsivity	EIS Impulsivity	EIS Venturesomeness	EIS Empathy	MPQ Stress Reaction	MPQ Alienation	MPQ Control	MPQ Harm Avoidance	STAI State Anxiety	STAI Trait Anxiety	BART duration (sec.)	BART pumps (ave.)	BART pumps (ave. low payoff)	BART pumps (ave. med. payoff)	BART pumps (ave. high payoff)	BART pumps adjusted ave.	BART pumps adj. ave.(low)	BART pumps adj. ave.(med)	BART pumps adj. ave.(high)	BART ave. payoff (cents)	BART balloon explosions (total)	BART explosions (low sched)	BART explosions (med)
Years flying																																								
Age	0.718																																							
Flight hrs tot	0.480																																							
FH past 12 mo			0.702																																						
FH past 90 d				0.753																																					
Hi perf hrs tot			0.524	0.616																																					
HP past 12 mo			0.631	0.685	0.499	0.850																																	
HP past 90 d			0.696			0.587	0.602																																
Simd instr hr tot			0.768	0.613	0.575	0.532	0.682	0.791																															
SI past 12 mo									0.479																														
SI past 90 d					0.516				0.481	0.841																													
Act inst hr tot			0.914	0.714	0.615		0.616	0.727	0.873																														
AI past 12 mo				0.730	0.808		0.487		0.618	0.561	0.625	0.668																											
AI past 90 d			0.499	0.662	0.727				0.501		0.539	0.696	0.873																										
ASA Aviation Safety Attitude Scale		-0.535				0.511																																	
ASI Anxiety Sensitivity Index,																																								
BIS Impulsive Planning																																								
BIS Motor Impulsivity																																								
BIS Cognitive Impulsivity																0.488	0.676																						
Barratt Impulsiveness Scale total																0.753		0.697																					
Eysenck Impulsivity Scale Impulsivity							0.478									0.660	0.515																						
EIS Venturesomeness		-0.533		0.469																																			
EIS Empathy																																								
HEI Hazardous Events Index	0.566		0.494																																				
MPQ Wellbeing																																								
MPQ Social Potency		-0.464												0.537																									
MPQ Achievement			0.508				0.504					0.483																											
MPQ Social Closeness																																								
MPQ Stress Reaction															0.468																								
MPQ Alienation																																								
MPQ Aggression															0.485						0.476	0.472																	
MPQ Control				-0.576												-0.648	-0.470		-0.782																				
MPQ Harm Avoidance														-0.464						-0.621																			
MPQ Traditionalism																																								
MPQ Absorption																																								
Rohrmann Cautiousness																						0.473		0.676															
Rohrmann Risk Propensity																				0.530																			
SSS Sensation Seeking Scale																	0.522			0.682				-0.490	-0.565														
STAI State Anxiety															0.484			0.470				0.496																	
STAI Trait Anxiety															0.561			0.485				0.661	0.531			0.740													
BART duration (seconds)																																								
BART pumps (average)																			-0.466									0.463											
BART pumps (ave. low paysched)																												0.467	0.825										
BART pumps (ave. med. payoff)																													0.934	0.648									
BART pumps (ave. high payoff)																													0.875	0.494	0.829								
BART pumps adjusted ave.																			-0.510									0.487	0.981	0.778	0.930	0.880							
BART pumps adjusted ave.(low)																												0.515	0.800	0.959	0.643	0.476	0.786						
BART pumps adjusted ave.(med)																													0.889	0.579	0.956	0.825	0.924	0.582					
BART pumps adjusted ave.(high)																			-0.497										0.842		0.809	0.968	0.875		0.831				
BART ave. payoff earned (cents)																																								
BART ave. payoff (low schedule)							-0.483																					0.489	0.498	0.502	0.471		0.544	0.672			0.743			
BART ave. payoff (med)																																0.469					0.800			
BART ave. payoff (high)																																					0.750			
BART balloon explosions (total)																													0.919	0.815	0.859	0.741	0.890	0.737	0.843	0.721				
BART explosions (low sched)																													0.705	0.900	0.508		0.628	0.788				0.808		
BART explosions (med)																													0.793	0.692	0.835	0.564	0.791	0.653	0.803	0.585		0.899	0.630	
BART explosions (high)																													0.797		0.793	0.883	0.801		0.851	0.839		0.786		0.585

Intercorrelation matrix, Pearson R, all variables with significance *p* < .01, Low Financial Incentive group (N=30).

Correlations (N=30 in all cases)	Years flying	Flight hrs tot	FH past 12 mo	FH past 90 d	Hi perf hrs tot	HP past 12 mo	HP past 90 d	Simd instr hr tot	SI past 12 mo	Act inst hr tot	AI past 12 mo	BIS Impulsive Planning	BIS Motor Impulsivity	BIS Cognitive Impulsivity	Barratt Impulsiveness Scale total	EIS Impulsivity	EIS Venturesomeness	MPQ Stress Reaction	MPQ Control	STAI State Anxiety	BART pumps (ave.)	BART pumps (ave. low payoff)	BART pumps (ave. med. payoff)	BART pumps (ave. high payoff)	BART pumps adjusted ave.	BART pumps adj. ave.(low)	BART pumps adj. ave.(med)	BART pumps adj. ave.(high)	BART ave. payoff (cents)	BART balloon explosions (total)
Age	0.718																													
FH past 12 mo		0.702																												
FH past 90 d			0.753																											
HP past 12 mo			0.685		0.850																									
HP past 90 d		0.696																												
Simd instr hr tot		0.768				0.682	0.791																							
SI past 90 d									0.841																					
Act inst hr tot		0.914	0.714				0.727	0.873																						
AI past 12 mo			0.730	0.808						0.668																				
AI past 90 d			0.662	0.727						0.696	0.873																			
BIS Cognitive Impulsivity													0.676		0.697															
Barratt Impulsiveness Scale total												0.753		0.697																
Eysenck Impulsivity Scale Impulsivity												0.660																		
MPQ Control																-0.782														
Rohrmann Cautiousness																			0.676											
SSS Sensation Seeking Scale																	0.682													
STAI Trait Anxiety																		0.661		0.740										
BART pumps (ave. low payoff sched)																					0.825									
BART pumps (ave. med. payoff)																					0.934									
BART pumps (ave. high payoff)																					0.875		0.829							
BART pumps adjusted ave.																					0.981	0.778	0.930	0.880						
BART pumps adjusted ave.(low)																					0.800	0.959			0.786					
BART pumps adjusted ave.(med)																					0.889		0.956	0.825	0.924					
BART pumps adjusted ave.(high)																					0.842		0.809	0.968	0.875		0.831			
BART ave. payoff (low schedule)																										0.672			0.743	
BART ave. payoff (med)																													0.800	
BART ave. payoff (high)																													0.750	
BART balloon explosions (total)																					0.919	0.815	0.859	0.741	0.890	0.737	0.843	0.721		
BART balloon explosions (low sched)																					0.705	0.900				0.788				0.808
BART balloon explosions (med)																					0.793	0.692	0.835		0.791	0.653	0.803			0.899
BART balloon explosions (high)																					0.797		0.793	0.883	0.801		0.851	0.839		0.786

These Rs reflect the relatively young median age of the pilots

Intra-BIS correlations

Intra-BART correlations

Intercorrelation matrix, Pearson R, all variables with significance $p < .0001$ (equivalent to $R^2 > .44$), Low Financial Incentive group (N=30). Note that most of the highly significant correlations have very simple explanations

BIS Impulsive Planning	EIS Impulsivity	EIS Venturesomeness	MPQ Stress Reaction	MPQ Control	STAI State Anxiety	**Non-Evident Correlations** where $R^2 > .44$ (N=30 in all cases)
0.660						Eysenck Impulsivity Scale Impulsivity
	-0.782					MPQ Control
				0.676		Rohrmann Cautiousness
		0.682				Sensation Seeking Scale
			0.661		0.740	STAI Trait Anxiety

Pearson Rs, variables with significance $p < .0001$ (equivalent to $.44 \leq R^2 \leq .61$) whose explanation is not obvious simply because they are correlated by their very nature (e.g. the various measures calculated from BART). The upshot here is that a) Each of these correlations is perfectly logical, and; b) Even this small number of correlations involves less than half the variance. That means that each instrument presumably measured different factors for the most fact, which was as it should be.

APPENDIX H

Description	Name	Sig.	g	MVs	Description	Name	Sig.
Order in which S. was run	runorder	0.982			Aviation Safety Attitude scale	asa	0.651
Subject ID#	**idnum**				Anxiety Sensitivity Index, total score	asi	0.143
Takeoff (Y/N)	**takeoff**				Barratt Impulsiveness Scale (BIS-10)--Impulsive Planning scale	imp_plan	0.902
Latency (minutes)	**latcy**				BIS--Motor Impulsivity scale	imp_motr	0.960
Visibility	vis	0.113			BIS-Cognitive Impulsivity scale	imp_cog	0.886
Ceiling	ceil	0.433			BIS--total score	bis	0.896
Incentive	incent	0.070			Eysenck Impulsivity Scale (EIS)--Impulsiveness scale	impuls	0.705
Experimenter	exptr	0.593			EIS--Venturesomeness scale	ventur	0.088
Yr started flying	yr_flg	0.785			EIS--Empathy scale	empath	0.277
Years flying, total	yrs_flg	0.966	2		Hazardous Events Index	hei	0.560
	age	0.653	1		Multidimensional Personality Questionnaire, Brief Form (MPQ-BF)		
	gender	0.461			MPQ--Wellbeing scale	wellbe	0.870
Type of License	type_lic	0.612			MPQ--Social Potency scale	socpot	0.468
Instrument Rating	inst_rtg	0.193			MPQ--Achievement scale	achieve	0.492
Total flight hours	fh_tot	0.410	2		MPQ--Social Closeness scale	socclose	0.290
Flthrs past year	fh_12m	0.536	2		MPQ--Stress Reaction scale	stress	0.528
Flthrs past 90 days	fh_90d	0.444	1		MPQ--Alienation scale	alienate	0.677
High-performance A/C, tot hrs	hp_tot	0.476	2		MPQ--Aggression scale	aggress	0.248
HP last 12 mo	hp_12m	0.287	2		MPQ--Control scale	control	0.540
HP last 90 days	hp_90d	0.151	2		MPQ--Harm Avoidance scale	harmav	0.614
Simulated instrument hrs total	si_tot	0.440	2		MPQ--Traditionalism scale	tradit	0.657
Sim hr last 12 mo	si_12m	0.239	2		MPQ--Absorption scale	absorpt	0.879
Sim hr last 90 d	si_90d	0.235	2		Rohrmann Risk Orientation Questionnaire--Cautiousness scale	roq_c	0.868
Actual instrument hrs, total	ai_tot	0.467	2		Rohrmann Risk Orientation Questionnaire--Risk Propensity scale	roq_p	0.225
AI last 12 mo	ai_12m	0.776	2		Sensation-Seeking Scale	sss	0.886
AI last 90 d	ai_90d	0.868	2		State-Trait Anxiety Inventory--State	anx_st	0.853
Personal visibility minimum	vis_min	0.386	2	1	State-Trait Anxiety Inventory--Trait	anx_tr	0.736
Personal ceiling minimum	ceil_min	0.955	2	1	Balloon Analogue Risk Task (BART)--test duration	durn_sec	0.565
Do you adjust minima?	min_adj	0.398		3	BART--average pumps	pumpsav	0.335
Flown this route before?	fln_rt	0.427		3	BART--average pumps, low-payoff condition	pmpavglo	0.465
Distance through wx imp?	dthruwx	0.813		3	BART--average pumps, medium-payoff condition	pmpavgm	0.630
$ bonus motivating? (Hi Incent only)	buck_mot	0.071		**32**	BART--average pumps, high-payoff condition	pmpavghi	0.198
Passengers change TO willingness	pass_mot	0.837		4	BART--adjusted average	pmpadjav	0.373
More flt hrs change TO willingness	fhincmot	0.893		5	BART--adjusted average, low-pay condn	padjavlo	0.782
Mission-critical chg. w-ness?	missnmot	0.020		**13**	BART--adjusted average, med-pay condn	padjavme	0.868
Flown Malibu chg w-ness?	mal_sub	0.840		**10**	BART--adjusted average, high-pay condn	padjavhi	0.207
Being a simulator chg w-ness?	sim_mot	0.127		**9**	BART--total payoff (cents)	pay_tot	0.790
...more specifically (re prev Q)	simmotsb	0.138		**14**	BART--total payoff, low-pay cond'n	pay_low	0.979
How significant was the $ to you?	buck_sig	0.164		**9**	BART--total payoff, med-pay cond'n	pay_med	0.630
Would crash embarrass you?	crashsig	0.048		**20**	BART--total payoff, high-pay cond'n	pay_hi	0.304
How much did you consider traffic	tx_mot	0.919		**49**	BART--total balloon explosions	bang_tot	0.259
Ever had a bad wx experience?	badwx	0.318		5	BART--explosions, low-pay cond'n	bang_low	0.422
					BART--explosions, med-pay cond'n	bang_me	0.325
					BART--explosions, high-pay cond'n	bang_hi	0.305

Predictor significances for the full data set (N=60), showing the reliability (expressed by the Wald *p*-value) of individual-predictor models (plus Constant) in logistic regression analysis with *Takeoff* as the dependent variable. Here each model included just one predictor, plus a constant. Subject ID is an identifier, not a predictor, and Latency is a descriptor, hence these lack *p*-values.

These predictors generally show very low reliability, with the exception of those highlighted in gray. However, of those, we should exclude all but *incent* and *ventur* from further consideration, due to high numbers of missing values (MV) for the other three. Note that the reference category for *lic_type* was "Private" (N=39), so *p* expresses the analysis "Private versus All Other Categories." No individual category had a *p* of < .12 in any case.

APPENDIX I

Description	Name	Sig.	g	MVs	Description	Name	Sig.
Order in w hich S. w as run	runorder	0.675			Aviation Safety Attitude scale	asa	0.645
Subject ID#	**idnum**				Anxiety Sensitivity Index, total score	asi	0.127
Takeoff (Y/N)	**takeoff**				Barratt Impulsiveness Scale (BIS-10)--Impulsive Planning scale	imp_plan	0.615
Latency (minutes)	**latcy**				BIS--Motor Impulsivity scale	imp_motr	0.957
Visibility	vis	0.064			BIS-Cognitive Impulsivity scale	imp_cog	0.398
Ceiling	ceil	0.691			BIS--total score	bis	0.562
Incentive	incent				Eysenck Impulsivity Scale (EIS)--Impulsiveness scale	impuls	0.394
Experimenter	exptr	0.261			EIS--Venturesomeness scale	ventur	0.713
Yr started flying	yr_flg				EIS--Empathy scale	empath	0.881
Years flying, total	yrs_flg	0.470	2		Hazardous Events Index	hei	0.221
	age	0.942	1		Multidimensional Personality Questionnaire, Brief Form (MPQ-BF)		
	gender	0.815			MPQ--Wellbeing scale	w ellbe	0.896
Type of License	type_lic	0.999			MPQ--Social Potency scale	socpot	0.269
Instrument Rating	inst_rtg	0.873			MPQ--Achievement scale	achieve	0.574
Total flight hours	fh_tot	0.591	1		MPQ--Social Closeness scale	socclose	0.590
Flthrs past year	fh_12m	0.911	2		MPQ--Stress Reaction scale	stress	0.544
Flthrs past 90 days	fh_90d	0.907			MPQ--Alienation scale	alienate	0.787
High-performance A/C, tot hrs	hp_tot	0.347	2		MPQ--Aggression scale	aggress	0.673
HP last 12 mo	hp_12m	0.713	2		MPQ--Control scale	control	0.930
HP last 90 days	hp_90d	0.328	2		MPQ--Harm Avoidance scale	harmav	0.641
Simulated instrument hrs total	si_tot	0.995			MPQ--Traditionalism scale	tradit	0.203
Sim hr last 12 mo	si_12m	0.588	2		MPQ--Absorption scale	absorpt	0.961
Sim hr last 90 d	si_90d	0.982	2		Rohrmann Risk Orientation Questionnaire--Cautiousness scale	roq_c	0.345
Actual instrument hrs, total	ai_tot	0.482	2		Rohrmann Risk Orientation Questionnaire--Risk Propensity scale	roq_p	0.637
AI last 12 mo	ai_12m	0.753	1		Sensation-Seeking Scale	sss	0.888
AI last 90 d	ai_90d	0.512	1		State-Trait Anxiety Inventory--State	anx_st	0.484
Personal visibility minimum	vis_min	0.523			State-Trait Anxiety Inventory--Trait	anx_tr	0.393
Personal ceiling minimum	ceil_min	0.487	1		Balloon Analogue Risk Task (BART)--test duration	durn_se	0.864
Do you adjust minima?	min_adj	0.244			BART--average pumps	pumpsav	0.341
Flow n this route before?	fln_rt	0.265			BART--average pumps, low -payoff condition	pmpavgl	0.552
Distance through w x imp?	dthruw x	0.627			BART--average pumps, medium-payoff condition	pmpavgr	0.462
					BART--average pumps, high-payoff condition	pmpavgh	0.234
Passengers change TO w illingness	pass_mot	0.175		1	BART--adjusted average	pmpadja	0.460
More flt hrs change TO w illingness	fhincmot	0.204			BART--adjusted average, low -pay condn	padjavlo	0.975
Mission-critical chg. w -ness?	missnmot	0.024		**7**	BART--adjusted average, med-pay condn	padjavm	0.768
Flow n Malibu chg w -ness?	mal_sub	0.854		**4**	BART--adjusted average, high-pay condn	padjavhi	0.186
Being a simulator chg w -ness?	sim_mot	0.910		**3**	BART--total payoff (cents)	pay_tot	0.749
...more specifically (re prev Q)	simmotsb	0.408		**7**	BART--total payoff, low -pay cond'n	pay_low	0.836
					BART--total payoff, med-pay cond'n	pay_med	0.990
Would crash embarrass you?	crashsig	0.337		**12**	BART--total payoff, high-pay cond'n	pay_hi	0.365
How much did you consider traffic	tx_mot	0.422		**26**	BART--total balloon explosions	bang_tot	0.272
Ever had a bad w x experience?	badw x	0.472			BART--explosions, low -pay cond'n	bang_low	0.458
					BART--explosions, med-pay cond'n	bang_me	0.403
					BART--explosions, high-pay cond'n	bang_hi	0.229

Predictor significances for the Low-Incentive data set (N=30), showing the reliability (Wald *p*-value) of individual-predictor models (plus Constant) in logistic regression analysis with *Takeoff* as the dependent variable. The reference category on *type_lic* is “Private,” on *simmotsb* it is “Didn’t matter.” Keep in mind that the SPSS reference category is the one being weighted “0” in the logistic regression prediction equation.

APPENDIX J

Description	Name	Sig.	g	MVs	Description	Name	Sig.
Order in which S. was run	runorder	0.612			Aviation Safety Attitude scale	asa	0.700
Subject ID#	**idnum**				Anxiety Sensitivity Index, total score	asi	0.910
Takeoff (Y/N)	**takeoff**				Barratt Impulsiveness Scale (BIS-10)--Impulsive Planning scale	imp_plan	0.809
Latency (minutes)	**latcy**				BIS--Motor Impulsivity scale	imp_motr	0.820
Visibility	vis	0.655			BIS-Cognitive Impulsivity scale	imp_cog	0.595
Ceiling	ceil	0.466			BIS--total score	bis	0.829
Incentive	incent	N/A			Eysenck Impulsivity Scale (EIS)--Impulsiveness scale	impuls	0.668
Weather severity	wxsvrty	0.364			EIS--Venturesomeness scale	ventur	0.085
Experimenter	exptr	0.080			EIS--Empathy scale	empath	0.296
Years flying, total	yrs_flg	0.451			Hazardous Events Index	hei	0.976
	age	0.420	1		Multidimensional Personality Questionnaire, Brief Form (MPQ-BF)		
	gender	0.476			MPQ--Wellbeing scale	wellbe	0.980
Type of License	type_lic	0.933			MPQ--Social Potency scale	socpot	0.947
Instrument Rating	inst_rtg	0.069			MPQ--Achievement scale	achieve	0.735
Total flight hours	fh_tot	0.420	2		MPQ--Social Closeness scale	socclose	0.117
Flthrs past year	fh_12m	0.385	1		MPQ--Stress Reaction scale	stress	0.980
Flthrs past 90 days	fh_90d	0.192			MPQ--Alienation scale	alienate	0.304
High-performance A/C, tot hrs	hp_tot	0.333			MPQ--Aggression scale	aggress	0.267
HP last 12 mo	hp_12m	0.090	2		MPQ--Control scale	control	0.622
HP last 90 days	hp_90d	0.164	2		MPQ--Harm Avoidance scale	harmav	0.337
Simulated instrument hrs total	si_tot	0.105			MPQ--Traditionalism scale	tradit	0.079
Sim hr last 12 mo	si_12m	0.036			MPQ--Absorption scale	absorpt	0.823
Sim hr last 90 d	si_90d	0.130	1		Rohrmann Risk Orientation Questionnaire--Cautiousness scale	roq_c	0.240
Actual instrument hrs, total	ai_tot	0.625	1		Rohrmann Risk Orientation Questionnaire--Risk Propensity scale	roq_p	0.325
AI last 12 mo	ai_12m	0.481	1		Sensation-Seeking Scale	sss	0.937
AI last 90 d	ai_90d	0.201	2		State-Trait Anxiety Inventory--State	anx_st	0.161
Personal visibility minimum	vis_min	0.519		1	State-Trait Anxiety Inventory--Trait	anx_tr	0.512
Personal ceiling minimum	ceil_min	0.726	1	1	Balloon Analogue Risk Task (BART)--test duration	durn_sec	0.437
Do you adjust minima?	min_adj	0.999		3	BART--average pumps	pumpsavg	0.703
Flown this route before?	fln_rt	0.485		3	BART--average pumps, low-payoff condition	pmpavglo	0.453
Distance through wx imp?	dthruwx	0.638		3	BART--average pumps, medium-payoff condition	pmpavgme	0.812
$ bonus motivating? (Hi Incent only)	buck_mot	0.071		2	BART--average pumps, high-payoff condition	pmpavghi	0.688
Passengers change TO willingness?	pass_mot	0.323		3	BART--adjusted average	pmpadjav	0.682
More flt hrs change TO willingness?	fhincmot	0.491		5	BART--adjusted average, low-pay condn	padjavlo	0.563
Mission-critical chg. w-ness?	missnmot	0.178		6	BART--adjusted average, med-pay condn	padjavme	0.835
Flown Malibu chg w-ness?	mal_sub	0.728		6	BART--adjusted average, high-pay condn	padjavhi	0.874
Being a simulator chg w-ness?	sim_mot	0.159		6	BART--total payoff (cents)	pay_tot	0.679
...more specifically (re prev Q)	simmotsb	0.139		7	BART--total payoff, low-pay cond'n	pay_low	0.868
How significant was the $ to you?	buck_sig	0.126		6	BART--total payoff, med-pay cond'n	pay_med	0.999
Would crash embarrass you?	crashsig	0.135		8	BART--total payoff, high-pay cond'n	pay_hi	0.995
How much did you consider traffic?	tx_mot	0.999		23	BART--total balloon explosions	bang_tot	0.503
Ever had a bad wx experience?	badwx	0.679		5	BART--explosions, low-pay cond'n	bang_low	0.482
					BART--explosions, med-pay cond'n	bang_med	0.533
					BART--explosions, high-pay cond'n	bang_hi	0.783

High Incentive data, N=30, single variable (plus Constant) models. Reference category for *type_lic* is “Private” (no individual $p < .187$). Reference category for *simmotsb* is “Positives and negatives cancel” (no individual $p < .072$).

APPENDIX K

Interactions *vis x ceil x factor below*								
Name	MV	Sig.	Ref		Sig.		Name	Sig.
runorder		0.189					imp_motr	0.381
							imp_cog	0.253
exptr		0.134	B	H	0.042		bis	0.392
yrs_flg		0.604					impuls	0.418
age		0.655					ventur	0.252
gender		0.942	M	F	0.444		empath	0.579
type_lic		0.788	Private				hei	0.466
inst_rtg		0.375	N	Y	0.131		w ellbe	0.482
fh_tot		0.282					socpot	0.512
fh_12m		0.559					achieve	0.606
fh_90d		0.473					socclose	0.938
hp_tot		0.380					stress	0.362
hp_12m		0.230					alienate	0.123
hp_90d		0.717					aggress	0.083
si_tot		0.244					control	0.772
si_12m		0.379					harmav	0.764
si_90d		0.802					tradit	0.054
ai_tot		0.305					absorpt	0.896
ai_12m		0.445					roq_c	0.201
ai_90d		0.177					roq_p	0.345
vis_min	1	0.308					sss	0.474
ceil_min	1	0.398					anx_st	0.179
min_adj	*3*	0.101	N	Y	0.997		anx_tr	0.235
fln_rt	*3*	0.363	N	Y	0.093		durn_sec	0.207
dthruwx	*3*	0.859	N	Y	0.961		pumpsavg	0.419
buck_mot	2	0.052	N	Y	0.137		pmpavglo	0.462
pass_mot	3	0.388					pmpavgme	0.568
fhincmot	5	0.194					pmpavghi	0.281
missnmot	6	0.291					pmpadjav	0.372
mal_sub	6	0.754					padjavlo	0.461
sim_mot	6	0.211					padjavme	0.519
simmotsb	7	0.316	"+/- cancel"				padjavhi	0.314
buck_sig	6	0.523					pay_tot	0.453
crashsig	8	0.227					pay_low	0.427
tx_mot	23	0.653					pay_med	0.568
badwx	5	0.806	N	Y	0.658		pay_hi	0.398
							bang_tot	0.438
asa		0.480					bang_low	0.553
asi		0.288					bang_med	0.571
imp_plan		0.643					bang_hi	0.278

High Incentive group only, N = 30. Reference category for *type_lic* was "Private," and results reflect composite significance for all license types. In no case was $p < .436$ for license type Reference category for *simmotsb* was "Positives and negatives cancelled." We were unable to coerce SPSS into defining the reference category as "Didn't matter." SPSS apparently sorts categoricals into frequency counts and assigns its "First" and "Last" categories according to frequency, rather than to the order in which categories are coded. In other words, recoding makes no difference. And, since its only options for assigning reference are "First" or "Last," it was impossible to equilibrate the analysis of *simmotsb* with its Low Incentive counterpart. In any event, the composite significance of *simmotsb* and all its components were all > .165, so the matter is irrelevant

APPENDIX L

Interactions *buck_mot x factor below*												
Name	MV	Sig.	Ref		Sig.		Name	MV	Sig.	Ref		Sig.
runorder	2	0.078	N	Y	0.144		asa	2	0.066	N	Y	0.061
vis	2	0.060	N	Y	0.125		asi	2	0.083	N	Y	0.187
ceil	2	0.064	N	Y	0.112		imp_plan	2	0.055	N	Y	0.053
exptr	2	0.064	NH	YH	0.916		imp_motr	2	0.049	N	Y	0.051
		0.877	NB	YB	0.068		imp_cog	2	0.048	N	Y	0.080
yrs_flg	2	0.531	N	Y	0.253		bis	2	0.041	N	Y	0.050
age	2	0.129	N	Y	0.083		impuls	2	0.020	N	Y	0.101
gender	2	0.469	NM	YM	0.706		ventur	2	0.021	N	Y	0.139
		0.030	NF	YF	0.071		empath	2	0.200	N	Y	0.120
type_lic	2	0.974	NP	YP	0.999		hei	2	0.301	N	Y	0.194
inst_rtg	2	0.811	NN	YN	0.066		wellbe	2	0.104	N	Y	0.135
		0.047	NY	YY	0.877		socpot	2	0.116	N	Y	0.167
fh_tot	2	0.231	N	Y	0.134		achieve	2	0.045	N	Y	0.069
fh_12m	2	0.670	N	Y	0.225		socclose	2	0.321	N	Y	0.061
fh_90d	2	0.904	N	Y	0.253		stress	2	0.181	N	Y	0.227
hp_tot	2	0.897	N	Y	0.152		alienate	2	0.156	N	Y	0.527
hp_12m	2	0.860	N	Y	0.103		aggress	2	0.127	N	Y	0.568
hp_90d	2	0.649	N	Y	0.070		control	2	0.363	N	Y	0.248
si_tot	2	0.917	N	Y	0.113		harmav	2	0.801	N	Y	0.088
si_12m	2	0.200	N	Y	0.244		tradit	2	0.053	N	Y	0.411
si_90d	2	0.230	N	Y	0.728		absorpt	2	0.027	N	Y	0.067
ai_tot	2	0.315	N	Y	0.190		roq_c	2	0.084	N	Y	0.166
ai_12m	2	0.107	N	Y	0.246		roq_p	2	0.032	N	Y	0.058
ai_90d	2	0.156	N	Y	0.700		sss	2	0.064	N	Y	0.080
vis_min	2	0.084	N	Y	0.052		anx_st	2	0.022	N	Y	0.138
ceil_min	2	0.387	N	Y	0.137		anx_tr	2	0.050	N	Y	0.062
min_adj	3	0.007	NN	YN	0.228		durn_sec	2	0.122	N	Y	0.336
		0.999	NY	YY	0.999		pumpsavg	2	0.063	N	Y	0.044
fln_rt	3	0.030	NN	YN	0.999		pmpavglo	2	0.032	N	Y	0.039
		0.707	NY	YY	0.999		pmpavgme	2	0.095	N	Y	0.038
dthruwx	3	0.092	NN	YN	0.999		pmpavghi	2	0.095	N	Y	0.075
		0.684	NY	YY	0.825		pmpadjav	2	0.070	N	Y	0.044
pass_mot	4	0.169	N	Y	0.499		padjavlo	2	0.039	N	Y	0.029
fhincmot	5	0.256	N	Y	1.000		padjavme	2	0.132	N	Y	0.039
missnmot	6	0.477	N	Y	0.999		padjavhi	2	0.122	N	Y	0.083
mal_sub	6	0.945	N	Y	0.503		pay_tot	2	0.059	N	Y	0.043
sim_mot	6	0.251	N	Y	1.000		pay_low	2	0.042	N	Y	0.029
simmotsb	7	0.442	N	Y	0.168		pay_med	2	0.107	N	Y	0.042
buck_sig	6	0.031	N	Y	0.120		pay_hi	2	0.083	N	Y	0.071
crashsig	8	0.124	N	Y	0.387		bang_tot	2	0.080	N	Y	0.054
tx_mot	23	0.999	N	Y	0.999		bang_low	2	0.047	N	Y	0.052
badwx	5	0.840	NN	YN	0.414		bang_med	2	0.120	N	Y	0.082
		0.143	NY	YY	0.212		bang_hi	2	0.153	N	Y	0.077

High Incentive, N = 30. Reference category for *type_lic* is "Private." Reference category for *simmotsb* is "Positives and negatives cancel." The reason for most of the missing values here is that *buck_mot* had two itself, so each analysis therefore automatically had to reflect at least these two.