1 Summary

- There is no effect of the variance manipulation, overall.
- Just eyeballing the data, the Mechanical Turk data is extremely noisy, especially in the first block (Figure 3, top), and less so in the second block (middle). This might be due to delayed onset of adaptation which then stabilizes by the second block (Figure 4).
- There is a tendency for people to be less consistent at the /f/ (high) end of the continuum thatn the /s/ (low) end. This is visible both as more variability in ID responses between subjects at the /f/ end (Figure 1) and in more "errors" on the five most extreme stimuli on the continuum at each end (/s/ responses to the five most /f/-like and vice versa; Figure 10).
- While there is overall more variability in the MTurk ID data (suggesting that the variability in audio equipment used might particularly impact this phonetic continuum), a similar pattern (problem with /f/ end) also shows up in the lab data, with a slight correlation between slope and intercept (lower slopes appear to have maximally ambiguous stimulus shifted towards /f/ end; Figure 1).
- Relatedly, the degree of shift is much smaller for those subjects who make more "errors" during pre-test (ID) on the stimuli that they will see as adaptors later on (Figure 11).
- The variance manipulation seems to have different effects depending on where the subject's category boundary is, as measured by the maximally ambiguous stimulus during ID (Figure 8). This makes sense given that the category boundary is where large shifts in reponse proportion could, in principle, be observed. Some of this might be corrected by using continua coordinates centered around each subject's boundary (Figure 9), but this normalization don't seem to produce clear effects from the variance manipulation.
- We ran a follow-up experiment looking at whether subjects' boundaries will become shallower when exposed to high variance distributions of *both* /f/ and /s/, interleaved. If subjects are sensitive to the variance of the type manipulated in the adaptation experiment, then there should be some difference between exposure to high vs. low variance distributions of both categories. We did not find any such difference (Figure 15, etc.), which leads us to suspect that the assumption that subjects are even sensitive to variability in these stimuli (in the range we're using) doesn't hold up.
- There are a variety of reasons why we may have failed to find the expected effect. MTurk may not be appropriate for this design, given the level of between-subject variability we observe (although the failure to find any difference in boundary slope holds even when comparing pre-test with the post-tests). The meta-linguistic nature of the category-judgement task may somehow interfere with the tracking of variance statistics. The particular stimuli we're using might not show enough acoustic variability in the tested range, or the joint statistics of the whole suite of cues that are manipulated by the spectral blending may not result in a clean projection onto a one-dimensional continuum (as is assumed when treating variance in continuum index as variance in the listener's perceptual space).

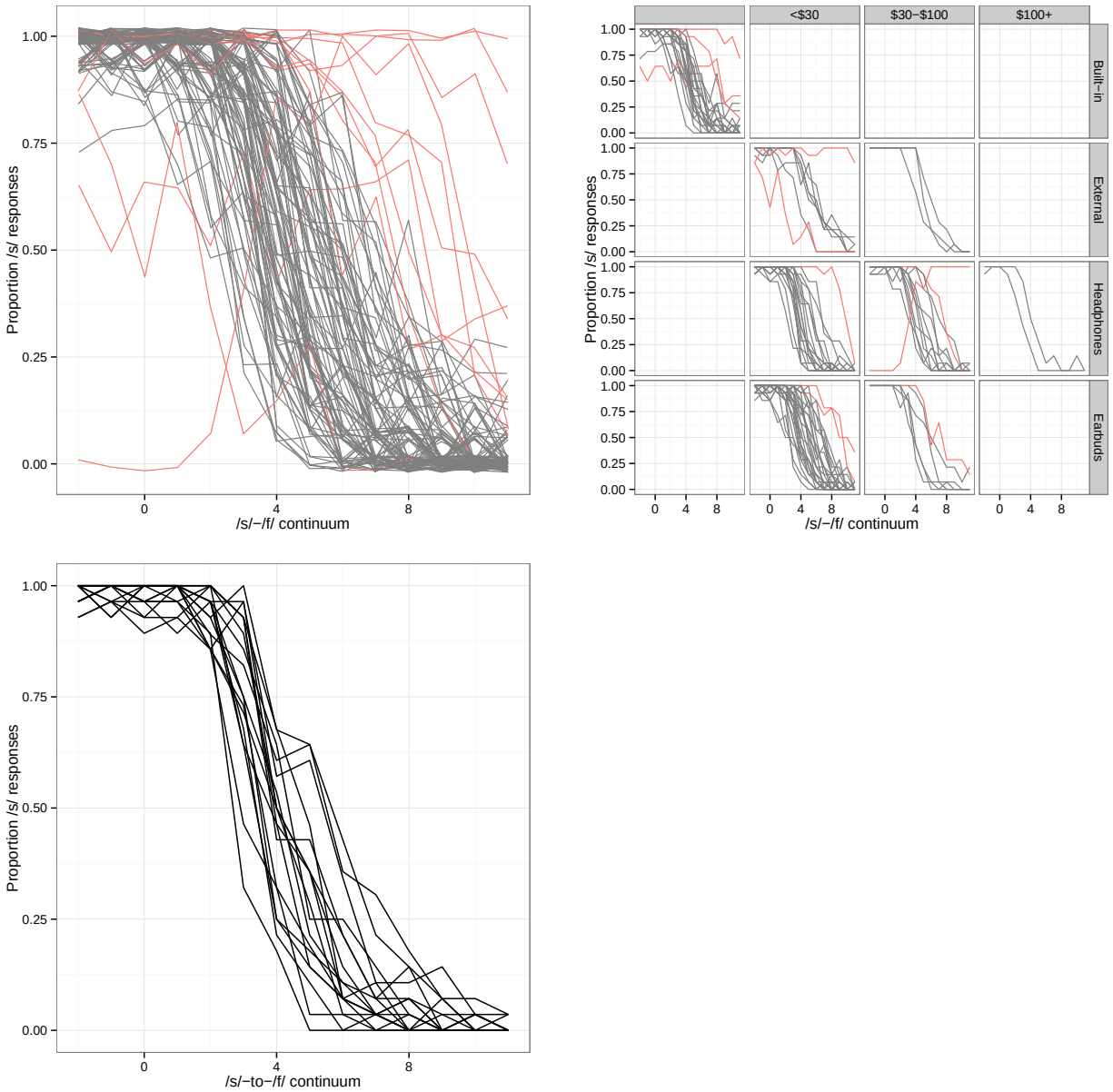

Figure 1: Pre-test, f/s classification data from Mechanical Turk subjects, plotted all together (left) and broken out by subject-reported audio equipment type (right). Lab data from Arty (bottom).

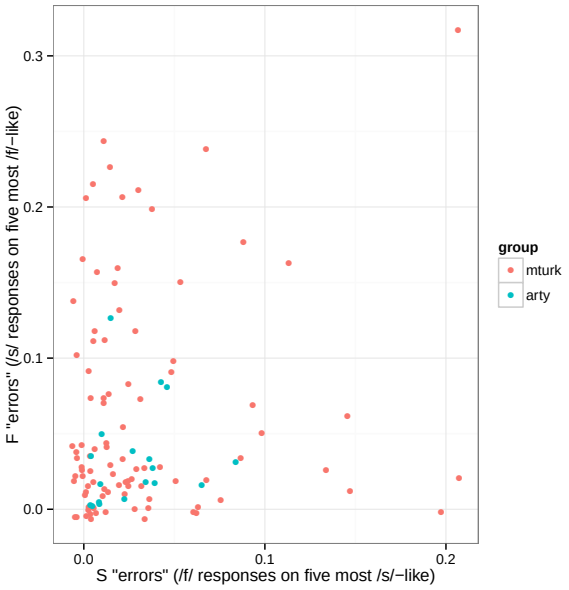

Figure 2: "Errors" in ID responses (for five most prototypical for each category, proportion of other-category responses). S errors (/f/ responses to the five most /s/-like continuum items) on the x-axis, and F errors on the y-axis (points are jittered for better visibility). Most subjects show rather low error rates for both /f/ and /s/, but for those subjects showing errors the /f/ error rate is generally higher (although this is mostly limited to the MTurk subjects). MTurk subjects who failed calibration are excluded.

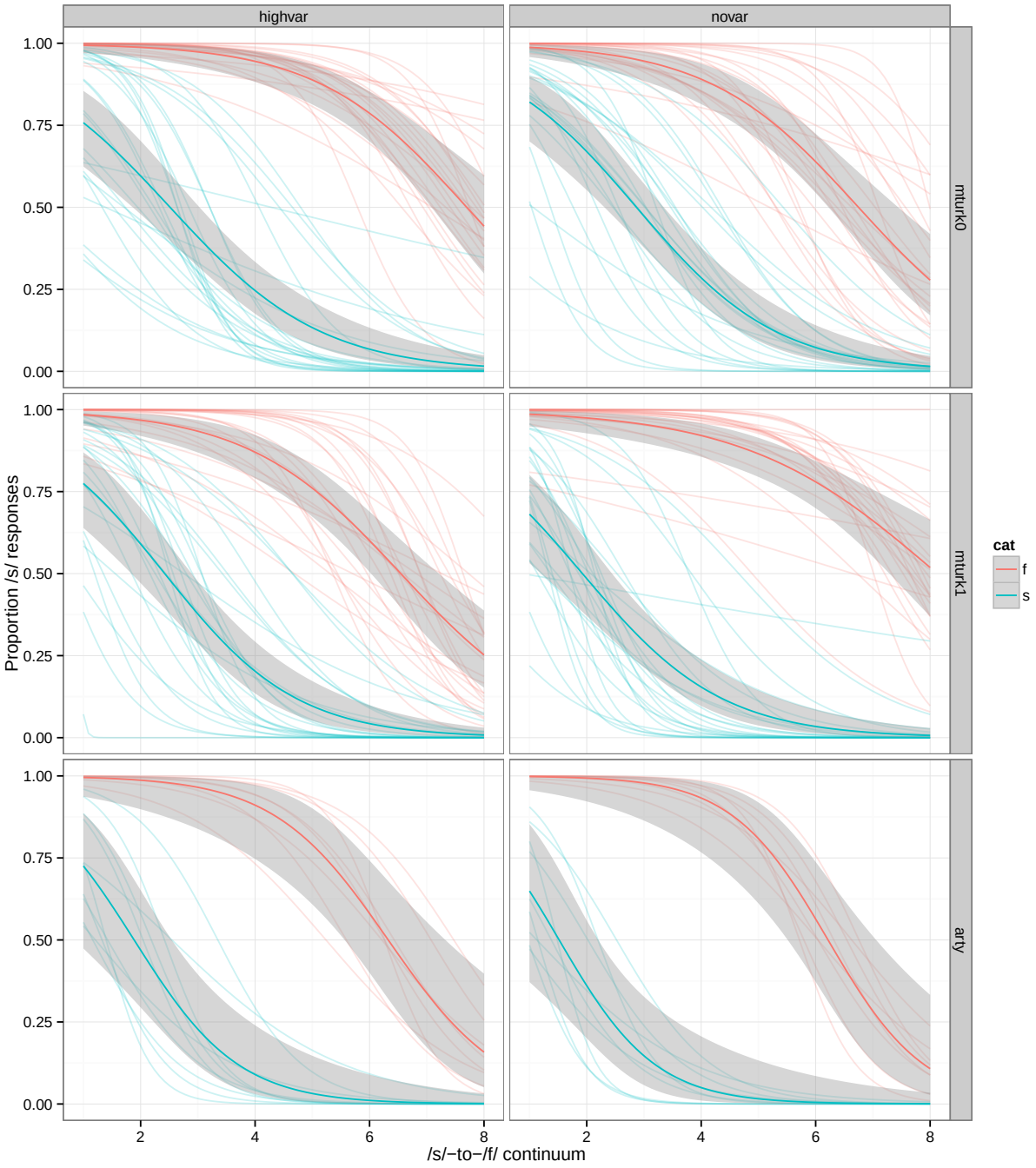

Figure 3: Shifted boundaries for lab and web data, averaged over entire adaptation session. Faint curves show (logistic smoothed) by-subject data, and thick lines show logistic fit to aggregate data. Each row shows one group: MTurk data, first block, MTurk data, second block, and lab (Arty) data, both blocks. First block MTurk data is extremely noisy.

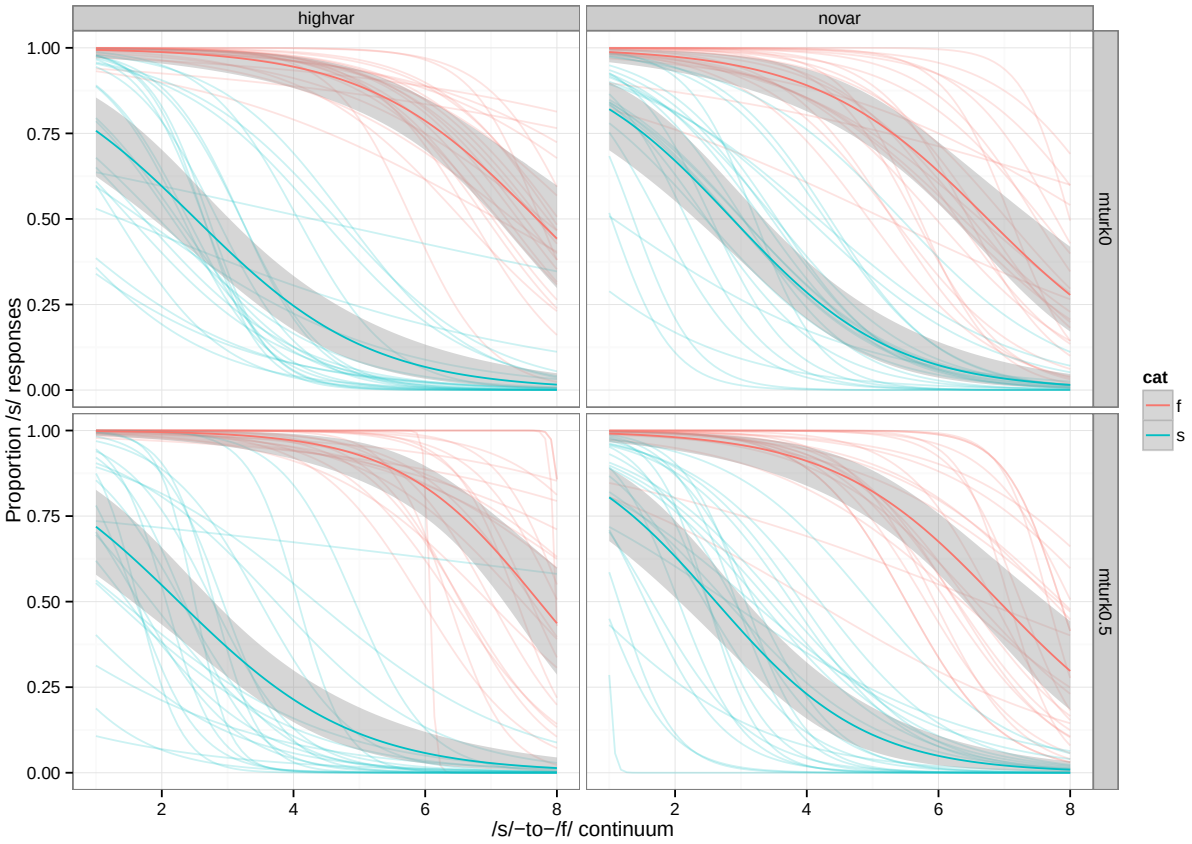

Figure 4: Shifted boundaries for MTurk data, comparing entire first block vs. only second half of first block.

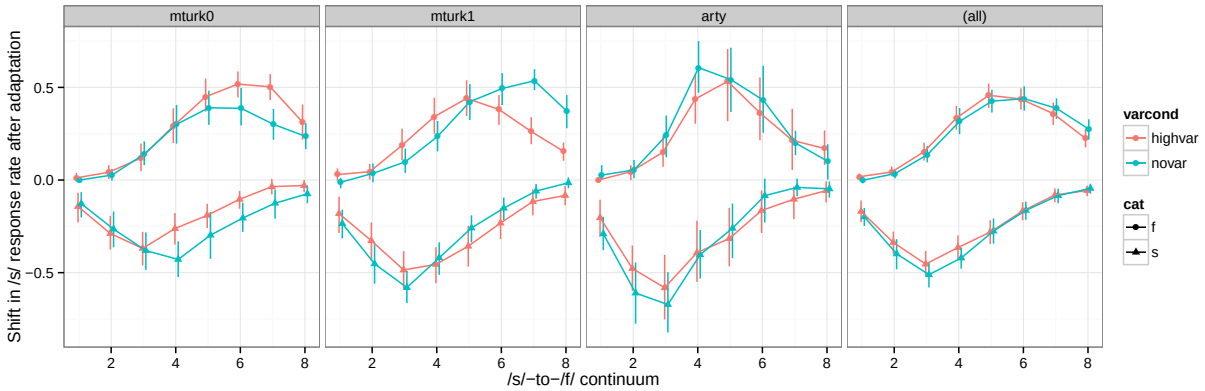

Figure 5: Mean shift in /s/ response rate after adaptation, by condition. Error bars show 95% confidence intervals on the mean.

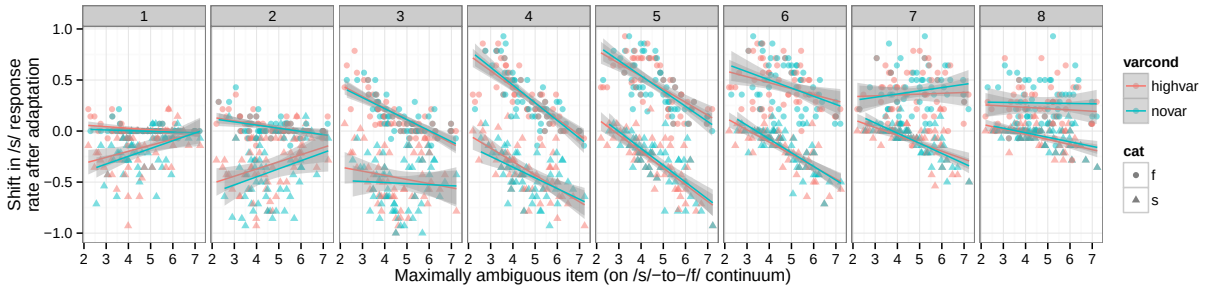

Figure 6: Effect of subject's maximally ambiguous stimulus on shift observed at each continuum location. The x-axis shows the maximally ambiguous item from pre-test, the y-axis shows the shift (change in /s/ responses after adaptation). Each facet shows one test stimulus (from /s/ to /f/ on the continuum). Circles are /f/ adaptors and triangles are /s/ adaptors, while red shows the high variance/range condition and blue the no variance/fixed condition. Linear smoothers are also shown. Overall there is a *negative* trend in the shifts with more /f/-like max-ambig, which is more pronounced when the shift is measured at intermediate positions (where shifts are generally larger). This means that subjects whose maximally ambiguous stimulus is more /f/-like show *stronger* (negative shift) adaptation effects after /f/ exposure, and *weaker* (positive shift) adaptation after /s/.

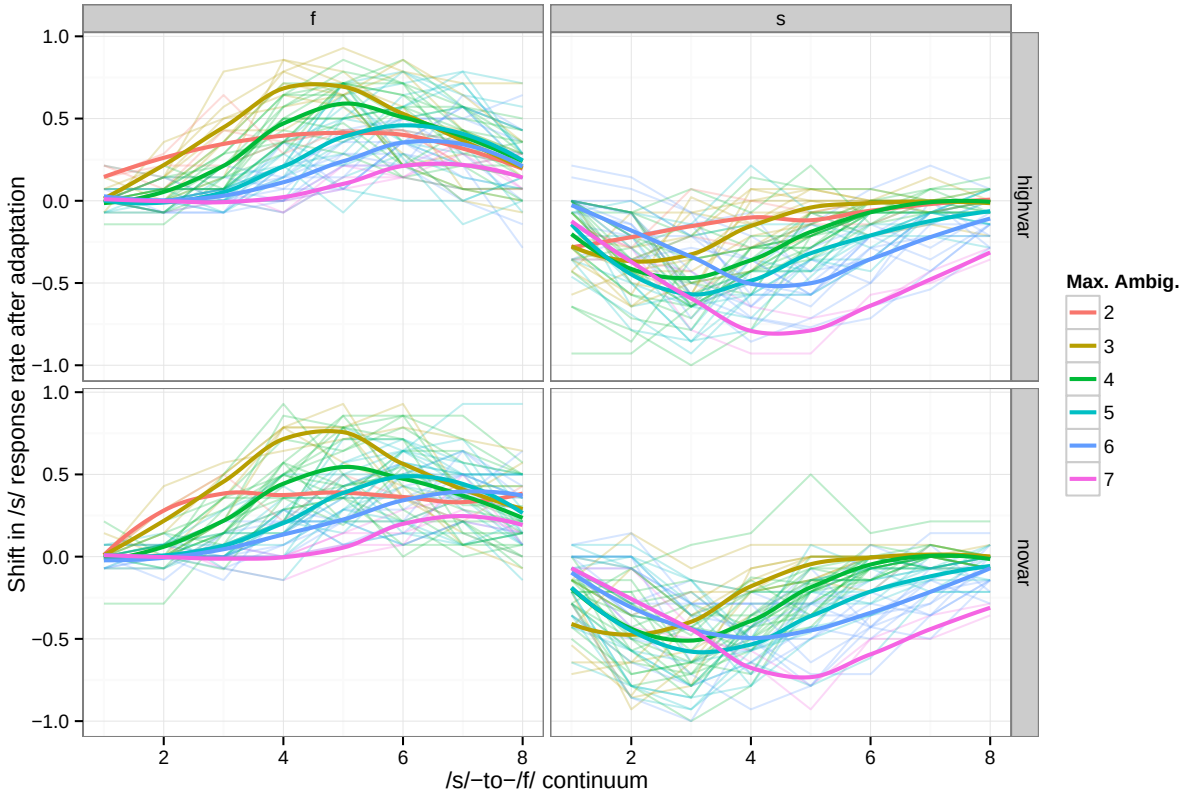

Figure 7: Individual subjects' shifts, as in Figure 5, colored by their maximally ambiguous stimulus during pretest.

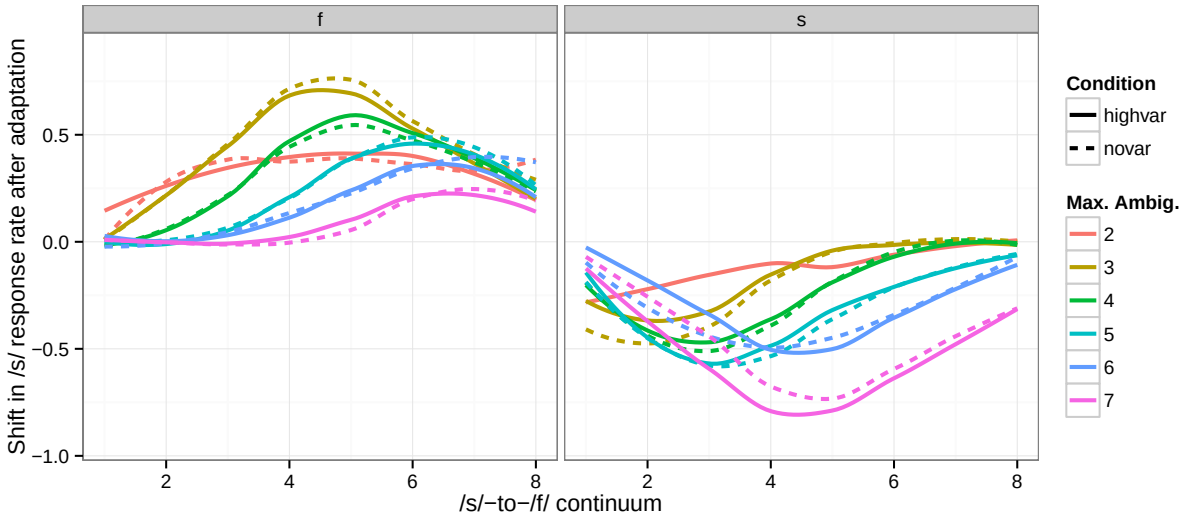

Figure 8: Effect of variance and adaptor category on shift, broken out by pre-test maximally ambiguous stimulus. Lines plotted are Loess smoothed averages, confidence intervals not shown.

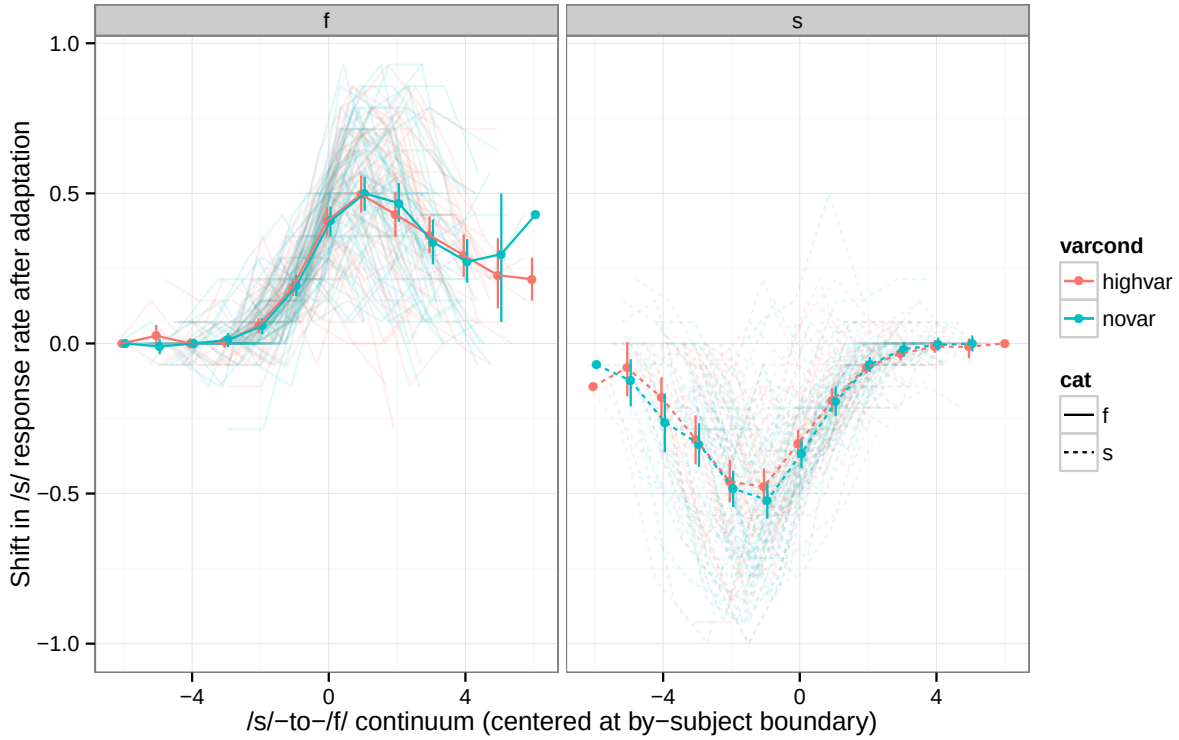

Figure 9: Shifts, with continuum centered around each subject's boundary (maximally ambiguous stimulus). Errorbars are bootstrapped 95% confidence intervals on the mean shift for the rounded, normalized continuum position. Centering reduces variance of shift from 0.04 to 0.035

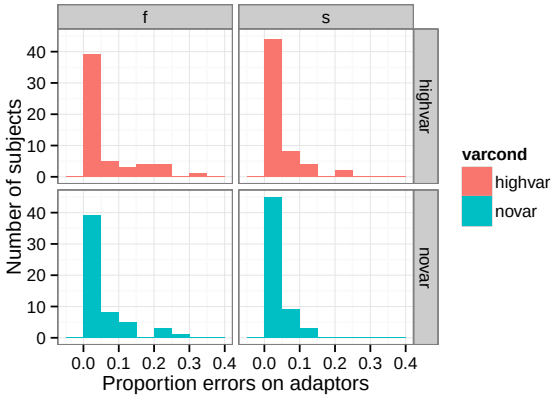

Figure 10: Distribution of adaptor ambiguity, as determined by the number of "errors" (proportion of other-category responses) during pretest to stimuli that were used as the adaptors in the next block ($1-5$ and $10-14$ for the high-variance /s/ and /f/, and 3 and 12 for the no-variance /s/ and /f/, respectively). The most notable trend is that there is a much longer tail of high error rates in the highvar/f condition, as Arty noted. In this condition there are 12 subjects with more than 10% errors.

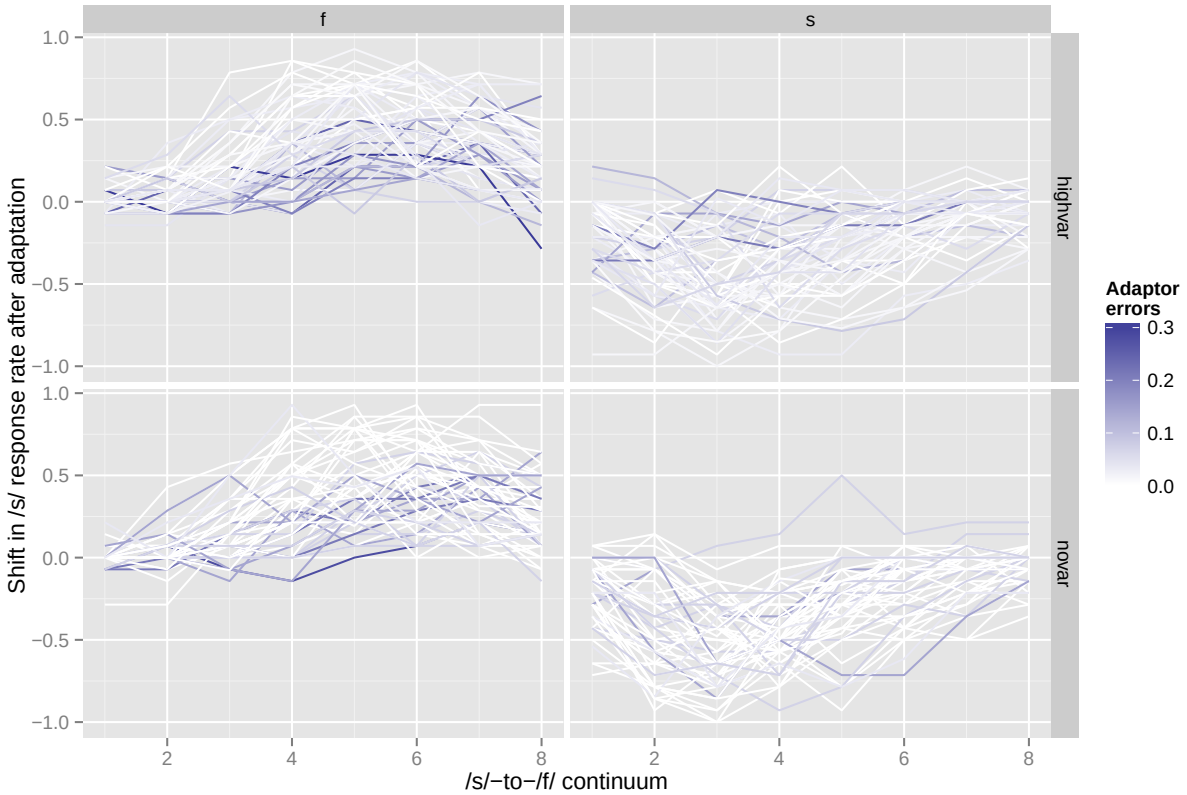

Figure 11: Shifts by subject, colored according to adaptor error rate (see Figure 10). As we suspected, people with higher error rates (more ambiguous percepts for adaptors) had smaller shifts, suggesting that the early effects we observed in the lab data might be accounted for by differences between the distribution of adaptor ambiguity across the conditions.

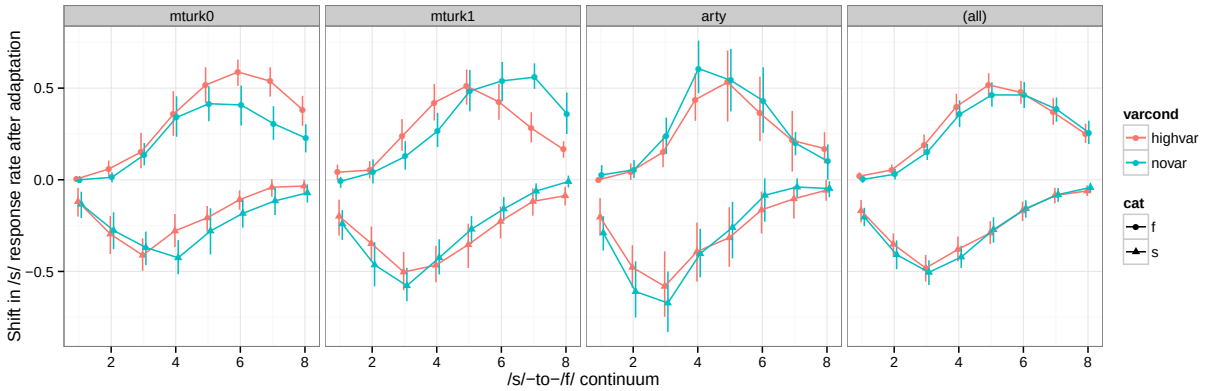

Figure 12: Mean shift after excluding subjects with adaptor ambiguity > 10% (compare to Figure 5).

```
>
> mu <- c(-4, 16)
>
> cal.a.reganal <- ddply(cal.a, .(subject, varcond, category), function(d) {
+     m <- glm(resp ~ poly(stim, 2, raw = T), family = "binomial", data = d)
+
+     A <- coef(m)[3]
+     B <- coef(m)[2]
+     C <- coef(m)[1]
+
+     # curve(1 / (1 + exp(-(A*x^2 + B*x + C))), from=mu[1], to=mu[2]) curve(1 /
+     # (1 + exp(B*x + C)), from=mu[1], to=mu[2], add=T)
+
+     # fit variances numerically
+     xs <- seq(mu[1], mu[2], by = 0.1)
+     optimized <- optim(c(1, 1), function(ls) {
+         mean((plogis(exp(-(A * xs^2 + B * xs + C))) - plogis(dnorm(xs, mu[2],
+             sqrt(1/ls[2]))/dnorm(xs, mu[1], sqrt(1/ls[1]))))^2)
+     })
+
+     ls.opt <- optimized$par
+
+     # curve(dnorm(x, mu[1], sqrt(1/ls.opt[1])) * sqrt(2*pi/ls.opt[1]), add=T,
+     # col='red') curve(dnorm(x, mu[2], sqrt(1/ls.opt[2])) *
+     # sqrt(2*pi/ls.opt[1]), add=T, col='red') curve(1 / (1 + dnorm(x, mu[2],
+     # sqrt(1/ls.opt[2]))/dnorm(x, mu[1], sqrt(1/ls.opt[1]))), add=T,
+     # col='red')
+
+     return(data.frame(var.s = 1/ls.opt[1], var.f = 1/ls.opt[2]))
+ })
>
> head(tca.wfits <- merge(cal.a.reganal, test.agg.cal))

##   subject varcond category var.s var.f cat stim.continuum blockn   resp
## 1    AANO highvar        s 16.67 44.47   s              2     NA 0.2857
## 2    AANO highvar        s 16.67 44.47   s              3     NA 0.1429
## 3    AANO highvar        s 16.67 44.47   s              8     NA 0.0000
## 4    AANO highvar        s 16.67 44.47   s              7     NA 0.0000
## 5    AANO highvar        s 16.67 44.47   s              1     NA 0.6429
## 6    AANO highvar        s 16.67 44.47   s              5     NA 0.0000
##       x  n group resp.cal   shift maxambig  slope stim.centered
## 1 4.000 14  arty   0.7857 -0.5000    5.067 0.7973      -3.06724
## 2 2.001 14  arty   1.0000 -0.8571    5.067 0.7973      -2.06724
## 3 0.000 14  arty   0.0000  0.0000    5.067 0.7973       2.93276
## 4 0.000 14  arty   0.0000  0.0000    5.067 0.7973       1.93276
## 5 9.001 14  arty   1.0000 -0.3571    5.067 0.7973      -4.06724
## 6 0.000 14  arty   0.4286 -0.4286    5.067 0.7973      -0.06724
##   adaptor.errs
## 1      0.04999
## 2      0.04999
## 3      0.04999
## 4      0.04999
## 5      0.04999
## 6      0.04999
```

```
>
> noise.var <- 0
> tca.wfits <- ddply(tca.wfits, .(subject, cat, varcond), function(d) {
+     means <- c(-4, 16)
+     vars <- c(d$var.s[1], d$var.f[1])
+     id.fcn.cal <- function(x) {
+         1/(1 + dnorm(x, means[2], sqrt(vars[2]))/dnorm(x, means[1], sqrt(vars[1])))
+     }
+     d$resp.cal.pred <- id.fcn.cal(d$stim.continuum)
+     if (d$cat[1] == "s") {
+         means[1] <- 0
+     } else {
+         means[2] <- 9
+     }
+     if (d$varcond[1] == "novar") {
+         vars[ifelse(d$cat[1] == "s", 1, 2)] <- 1/12 + noise.var
+     } else {
+         vars[ifelse(d$cat[1] == "s", 1, 2)] <- 5^2 * 1/12 + noise.var
+     }
+     id.fcn.shift <- function(x) {
+         1/(1 + dnorm(x, means[2], sqrt(vars[2]))/dnorm(x, means[1], sqrt(vars[1])))
+     }
+     d$resp.pred <- id.fcn.shift(d$stim.continuum)
+     d
+ })
>
> summary(tca.wfits)

##     subject       varcond    category     var.s           var.f
##  AANO   : 16   highvar:144   f:128    Min.   : 3.50   Min.   :11.3
##  AMOR   : 16   novar  :136   s:152    1st Qu.: 8.17   1st Qu.:18.9
##  AYEO   : 16                           Median :10.17   Median :24.6
##  CMAN   : 16                           Mean   :12.10   Mean   :28.2
##  DNAY   : 16                           3rd Qu.:15.14   3rd Qu.:37.2
##  DPIS   : 16                           Max.   :34.54   Max.   :61.7
##  (Other):184
##  cat     stim.continuum     blockn         resp             x
##  f:128   Min.   :1.00   Min.   : NA   Min.   :0.000   Min.   : 0.00
##  s:152   1st Qu.:2.75   1st Qu.: NA   1st Qu.:0.000   1st Qu.: 0.00
##          Median :4.50   Median : NA   Median :0.258   Median : 3.62
##          Mean   :4.50   Mean   :NaN   Mean   :0.421   Mean   : 5.89
##          3rd Qu.:6.25   3rd Qu.: NA   3rd Qu.:0.857   3rd Qu.:12.00
##          Max.   :8.00   Max.   : NA   Max.   :1.000   Max.   :14.00
##                         NA's   :280
##        n         group        resp.cal          shift
##  Min.   :14   mturk0:  0   Min.   :0.0000   Min.   :-1.0000
##  1st Qu.:14   mturk1:  0   1st Qu.:0.0714   1st Qu.:-0.2857
##  Median :14   arty  :280   Median :0.3928   Median : 0.0000
##  Mean   :14                Mean   :0.4624   Mean   :-0.0418
##  3rd Qu.:14                3rd Qu.:0.9286   3rd Qu.: 0.1552
##  Max.   :14                Max.   :1.0000   Max.   : 0.9286
##
##     maxambig        slope       stim.centered     adaptor.errs
##  Min.   :2.91   Min.   :0.689   Min.   :-4.694   Min.   :0.0000
```

```
##  1st Qu.:3.67   1st Qu.:0.968   1st Qu.:-1.694   1st Qu.:0.0000
##  Median :4.08   Median :1.225   Median : 0.306   Median :0.0143
##  Mean   :4.21   Mean   :1.196   Mean   : 0.294   Mean   :0.0247
##  3rd Qu.:4.71   3rd Qu.:1.406   3rd Qu.: 2.306   3rd Qu.:0.0357
##  Max.   :5.69   Max.   :1.921   Max.   : 5.093   Max.   :0.0857
##
##  resp.cal.pred      resp.pred
##  Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0482   1st Qu.:0.000
##  Median :0.4104   Median :0.343
##  Mean   :0.4560   Mean   :0.472
##  3rd Qu.:0.8694   3rd Qu.:1.000
##  Max.   :0.9991   Max.   :1.000
```

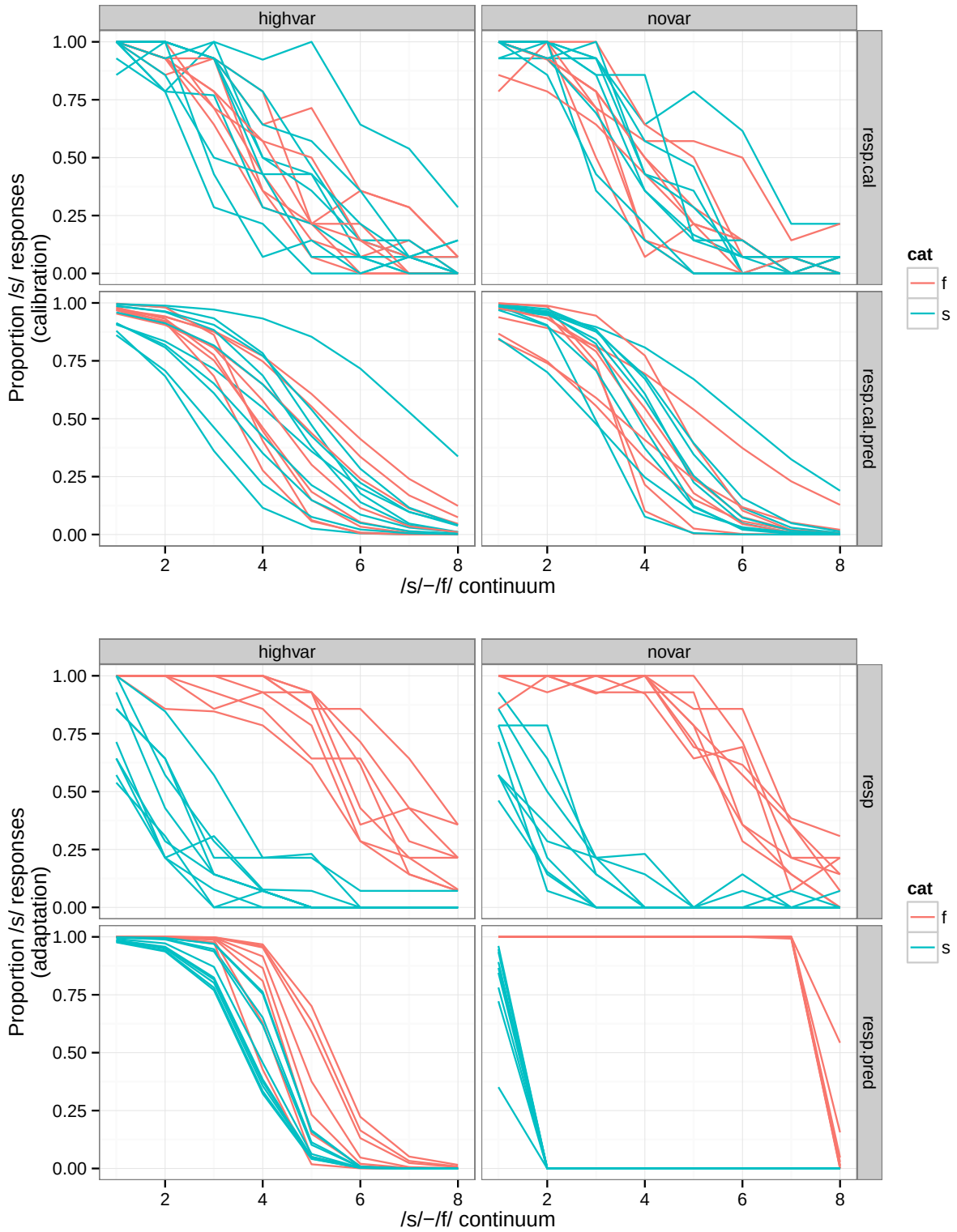

Figure 13: Predictions for calibration (bottom) and adaptation test (top), compared to actual data (top half of both) based on predicted assymptotic performance by belief-updating model and assuming that category means are located at true endpoints of continuum (which is two items beyond the /s/ and four beyond the /s/ end used in pre-test). The individual category variances were fit in order to fit each subject's pre-test performance (top panel). Predictions for post-adaptation performance were generated based on variance of 25.0833 for high-variance condition (variance of uniformly distributed random variable with range of five) and 0.0833 for no-variance condition (variance of uniformly distributed random variable with range of one), and shifted mean continuum positions of 0 for /s/ adaptors and 9 for /f/ adaptors. Updated categorization functions calculated assuming that the other category is unchanged.

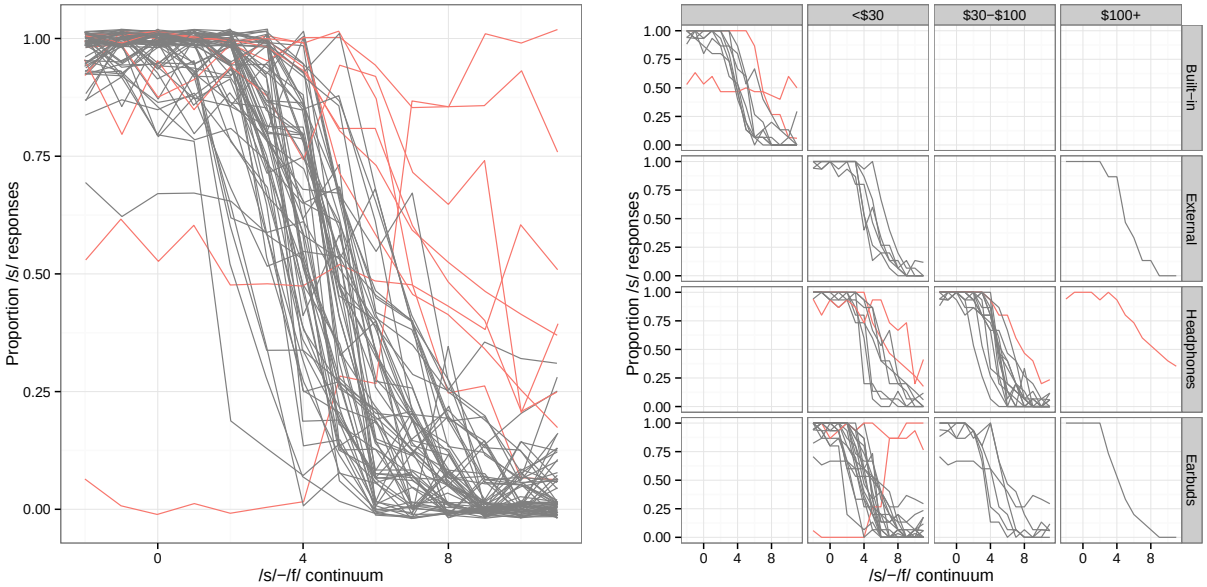

Figure 14: Pre-test (calibration) data from the mixed F/S presentation followup experiment.

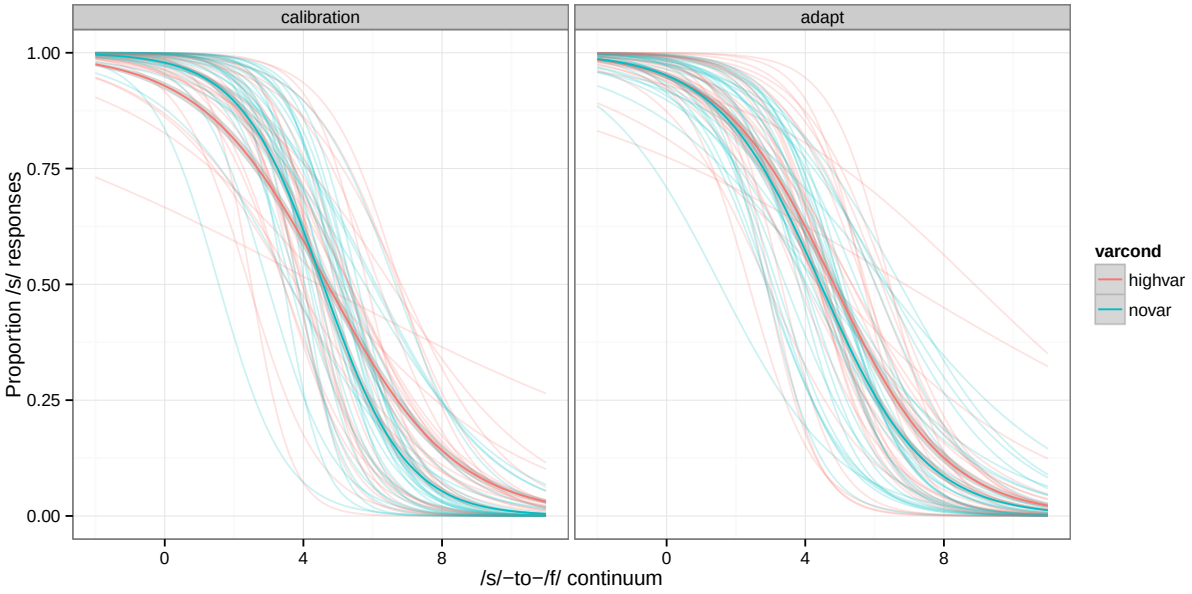

Figure 15: Categorization boundaries during calibration and during exposure for mixed f/s presentation

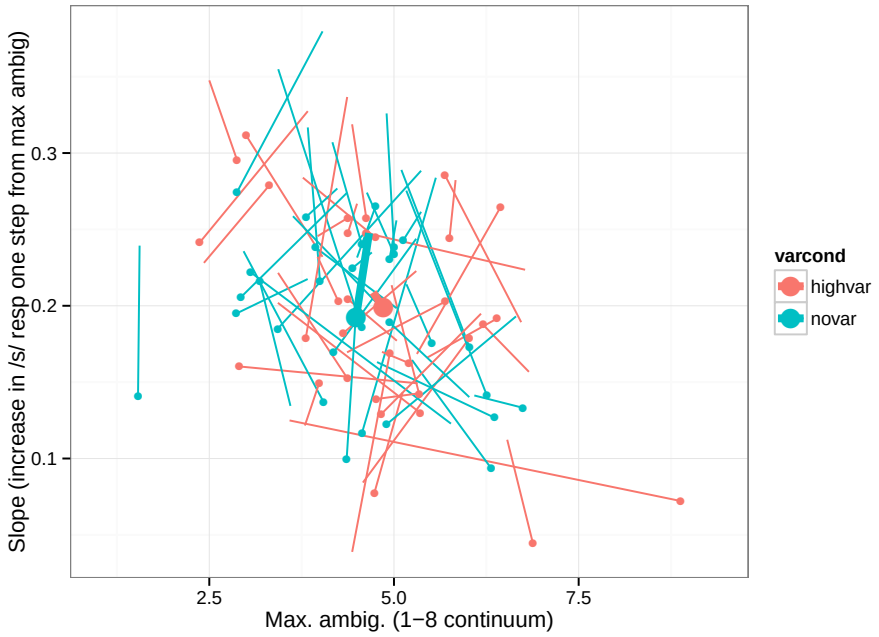

Figure 16: Visualization of results from mixed F/S presentation. Color shows variance condition. The "head" of each pin shows the slope and intercept of all test blocks, while the point shows where that subject started (pre-test). The big pins are the average across subjects, showing that there is no difference between the two conditions at test. There is a bit of a trend for slopes to become *shallower* in the no-variance condition (blue), but remain unchanged in the high-variance condition. Note that these measures (change in slope and intercept, averaged over subjects) is not the same as the change in the slope and intercept of the average *responses* (visualized in Figure 15), because of the non-linear transformation between the log-odds slope and intercept and the probability/binary responses.