Supplement to Interaction Effects on Common Measures of Sensitivity: Choice of Measure, Type I Error and Power

Stephen Rhodes, Nelson Cowan, Mario A. Parra, and Robert H. Logie

June 21, 2018

Contents

Summary

This document presents extra information to accompany the manuscript *'Interaction Effects on Common Measures of Sensitivity: Choice of Measure, Type I Error and Power'*. Plots of simulations not presented in the main manuscript are given. In addition, the raw output files used to create these plots are available at `https://github.com/stephenrho/MeasuresAndErrors/tree/master/results/tables` and can be read into `R` using the function `read.table`. Researchers wishing to test combinations of parameter settings not considered in the present work are able to do so using code available at `https://github.com/stephenrho/MeasuresAndErrors`. These materials are distributed under the terms of the GNU General Public License version 3.0 (`https://www.gnu.org/licenses/gpl-3.0-standalone.html`).

As described in the main manuscript, it has been shown that a misguided choice of sensitivity measure can lead to type I errors or reduced power for main effects, provided there are true group differences in bias (Schooler & Shiffrin, 2005; Rotello, Masson, & Verde, 2008). Here we assess the type I error rates and power of three commonly used sensitivity measures (d', P_r, and A') for *interactions* in data generated according to a signal detection or two-high threshold process. The main manuscript describes the simulations in full. To summarize, parameter settings were determined by the same linear equation, $p_{i,b,w} = \beta_0 + \beta_1 x_b + \beta_2 x_w + \beta_3 x_b x_w + b_i$. These linear effects are referred to throughout, with β_0 reflecting the grand mean, β_1 and β_2 reflecting the magnitude of main effects of group and condition, respectively, and β_3 controlling the size of the two-way interaction. The additional component, b_i, was also included to allow for random variability between simulated individuals. The model parameters to which these linear effects refer are indexed by superscripts throughout (d and k for the sensitivity and bias parameters in signal detection. P_r and B_r for two-high threshold). Further, the parameters are given on the axes of the figures presenting simulation results, along with a text description to aid interpretation of the figures (e.g. Mean sensitivity ($\beta_0^{(d)}$)).

Simulations 1–7 assess type I error, beginning with no main effects of group or condition on bias (1–3) and then introducing this variability (4–7). Following this up, simulations 8–12 look at the power of the three sensitivity measures under different circumstances. Simulations 8 and 9 do this without any variation in response bias and simulations 10–12 assess the influence of introducing bias differences on power. Questions regarding these simulations may be addressed to Stephen Rhodes (rhodessp@missouri.edu).

Type I Error Simulations

1 Simulation 1: varying overall sensitivity, main effects, and number of participants

In the first type I error simulations, we varied the mean sensitivity of observers, as well as the size of main effects of group and condition (fixed to be the same to reduce the number of simulations needed. Simulation 3 varies the two independently). In addition we varied the number of simulated subjects (12, 24, 48). The main manuscript reports the results of these simulations with the two-high threshold and equal variance signal detection models. Figures 1 and 2 present the results of unequal variance signal detection simulations. Note that $s = 0.8$ means the target distribution is more variable than the non-target distribution, whereas $s = 1.2$ has the opposite effect.

1.1 Unequal-variance signal detection

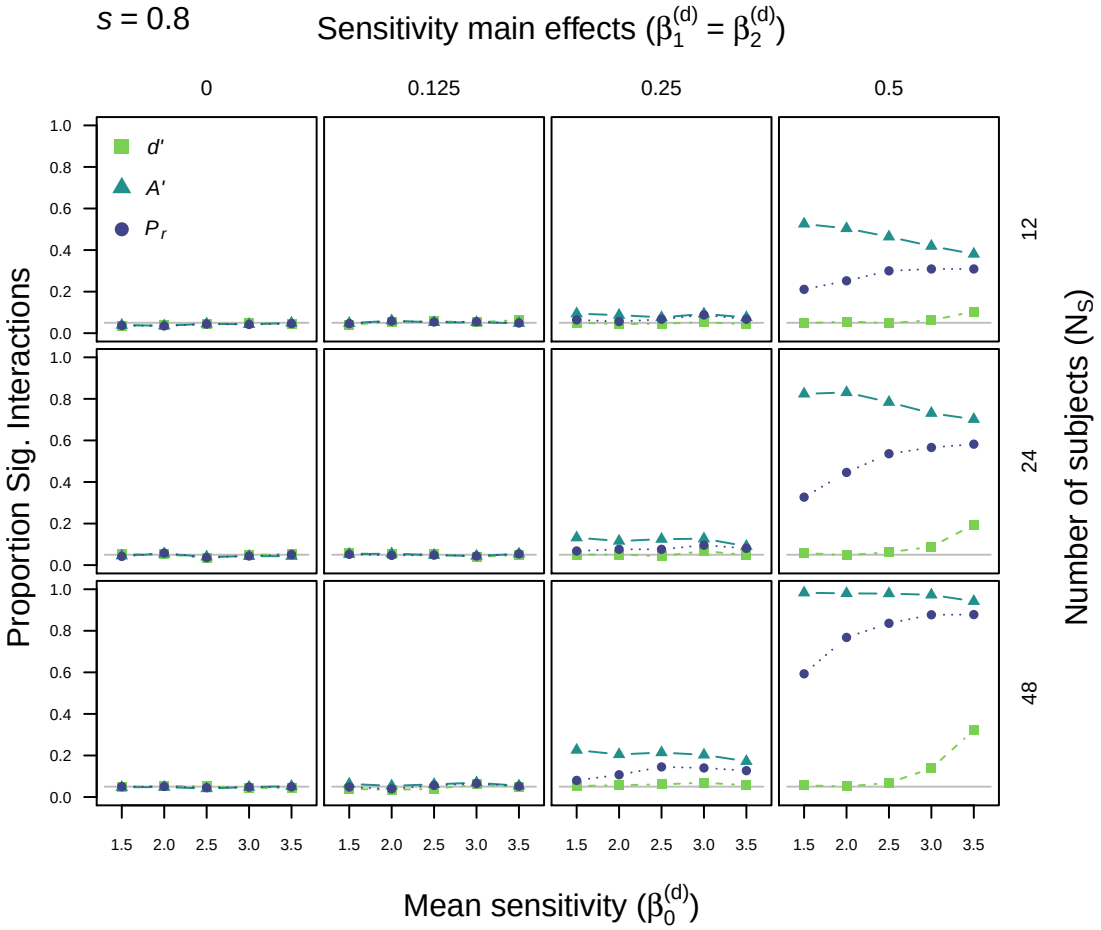

Figure 1: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects on sensitivity (left to right), and number of simulated participants (top to bottom).

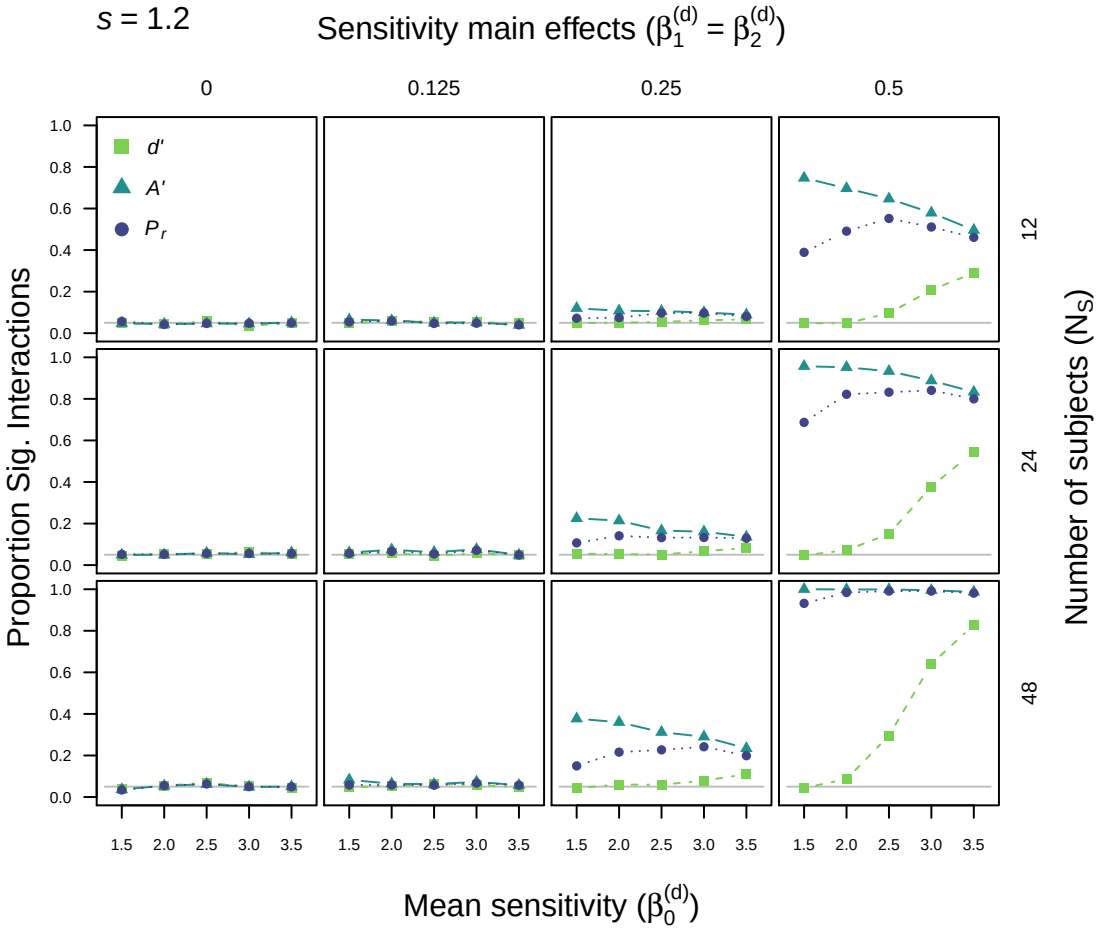

Figure 2: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects on sensitivity (left to right), and number of simulated participants (top to bottom).

2 Simulation 2: varying overall sensitivity, main effects, and number of trials

As noted in the main manuscript, varying the number of trials per condition led to largely similar type I error rates to varying the number of subjects per group. Figures 3 to 6 show these results.

2.1 Two-high threshold

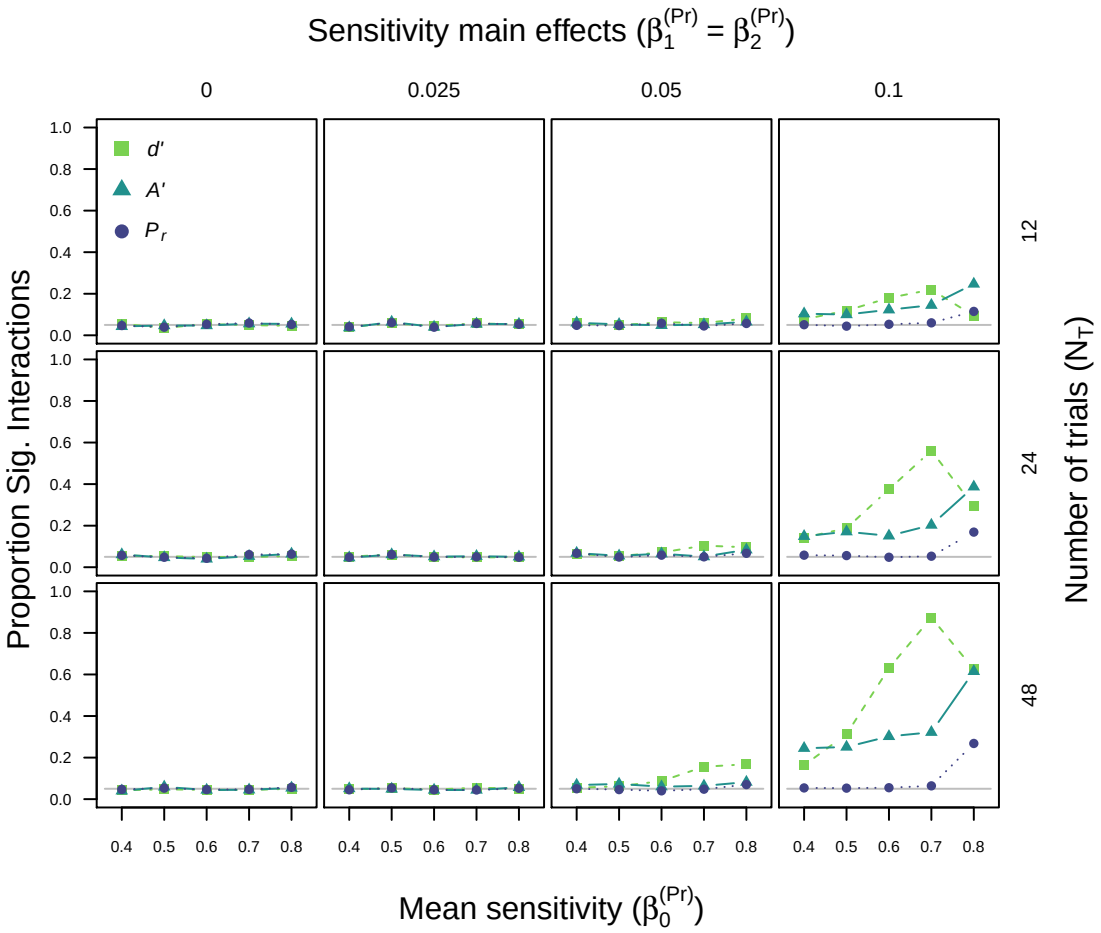

Figure 3: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall detection probability (x axis), magnitude of main effects on detection (left to right), and number of simulated trials (top to bottom).

2.2 Signal detection

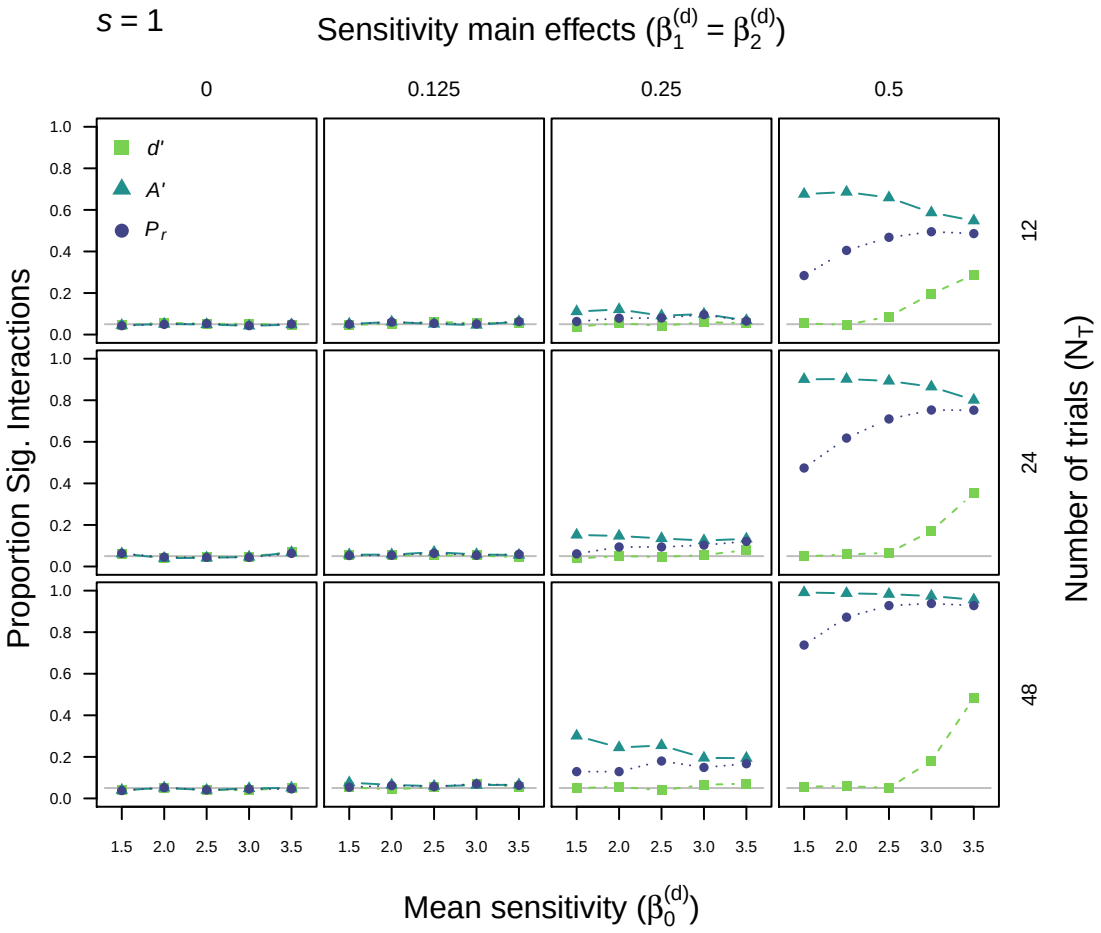

Figure 4: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall detection sensitivity (x axis), magnitude of main effects on sensitivity (left to right), and number of simulated trials (top to bottom).

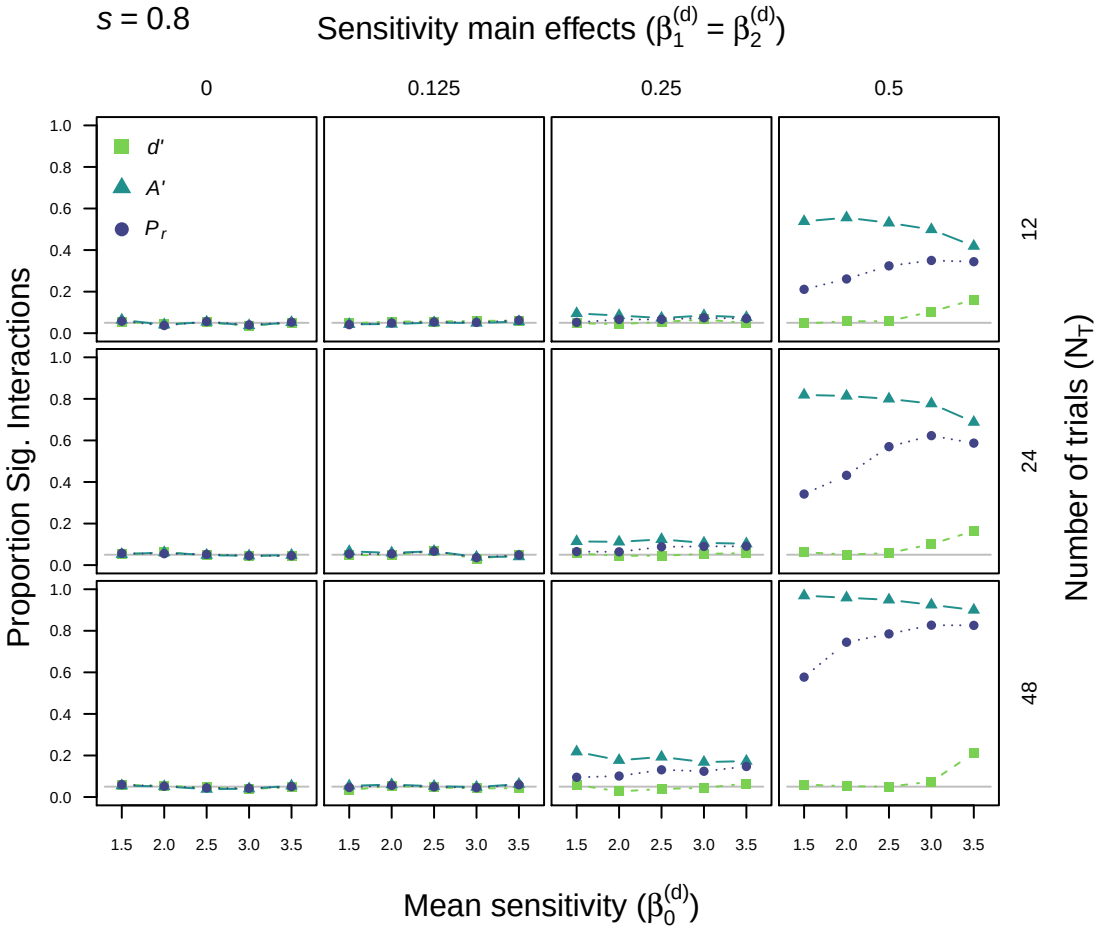

Figure 5: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects on sensitivity (left to right), and number of simulated trials (top to bottom).

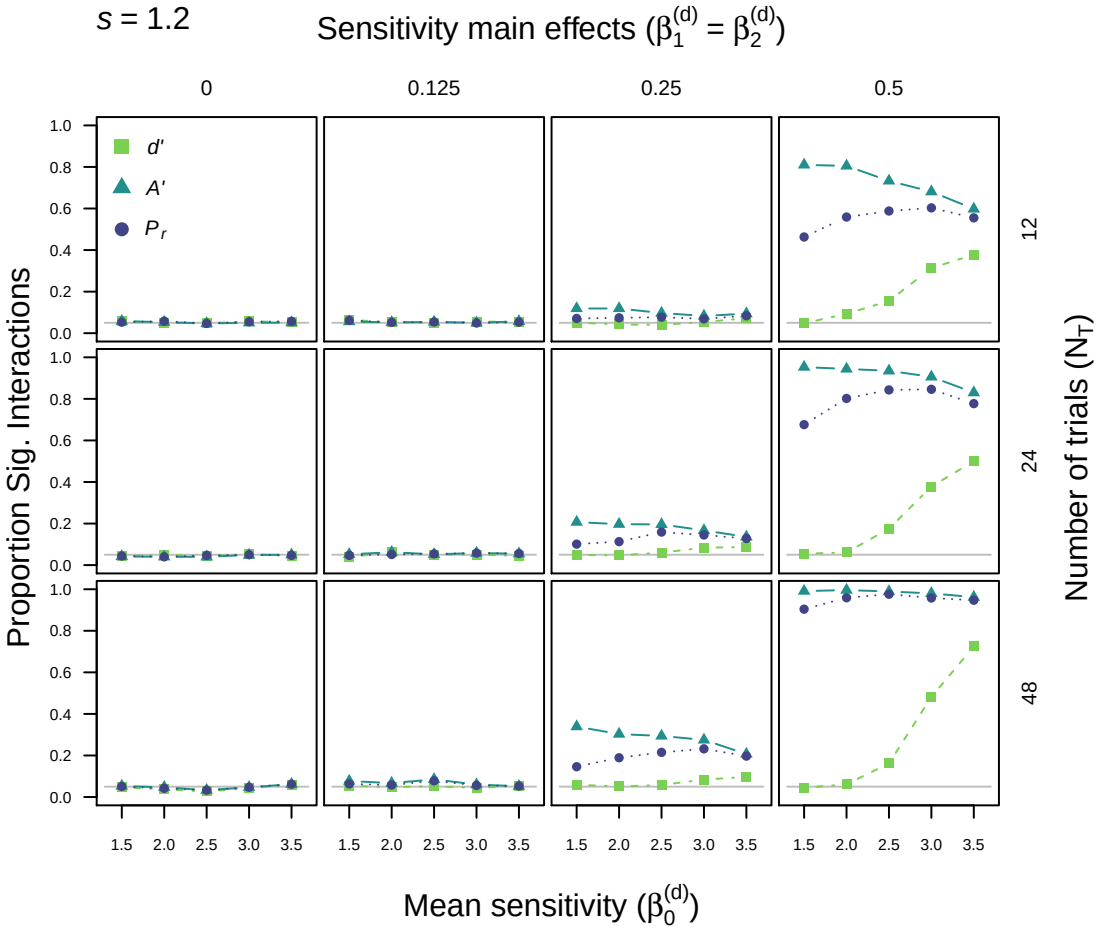

Figure 6: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects on sensitivity (left to right), and number of simulated trials (top to bottom).

3 Simulation 3: vary overall sensitivity and orthogonally vary main effects

Simulations 1 and 2 fixed the magnitude of the main effects of group and condition to be the same and suggested that, as long as both main effects are large, one obtains high type I error rates in the absence of variation in response bias. The third set of simulations orthogonally varied the size of the group and condition main effects to assess whether it is sufficient for only one main effect to be large to produce this finding. For these and the following type I error simulations the number of trials per condition and number of participants per group were both fixed to 24. Figures 7 to 10 show the outcome of these simulations for the two-high threshold and signal detection models.

3.1 Two-high threshold

For a high threshold generative model, error rates for d' approach 20% with one medium and one large main effect (see Figure 7). However, as shown by the initial simulations, the most pronounced error rates are observed with two-large main effects as seen in the bottom right hand panel of the figure.

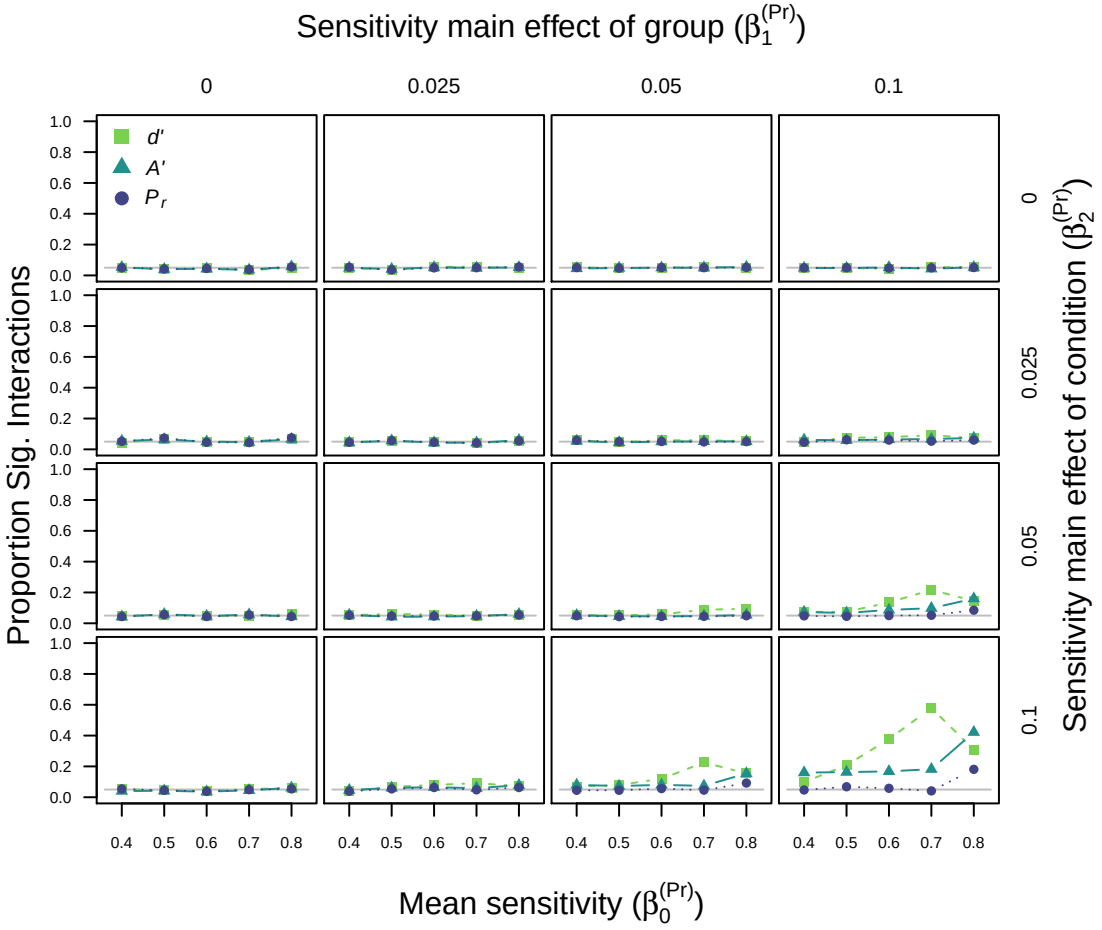

Figure 7: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall detection probability (x axis), magnitude of main effects of group (left to right) and condition (top to bottom) on detection (or sensitivity).

3.2 Signal detection

For a signal detection generative model, moderate error rates are observed when one effect is large and the other small (or when both are medium). The error rates become uncontrollable with one medium and one large effect, and rise to consistently over 50% for A' and P_r with two large main effects on sensitivity (Figure 8). This pattern is somewhat attenuated for $s = 0.8$ (Figure 9) and exacerbated when $s = 1.2$ (Figure 10).

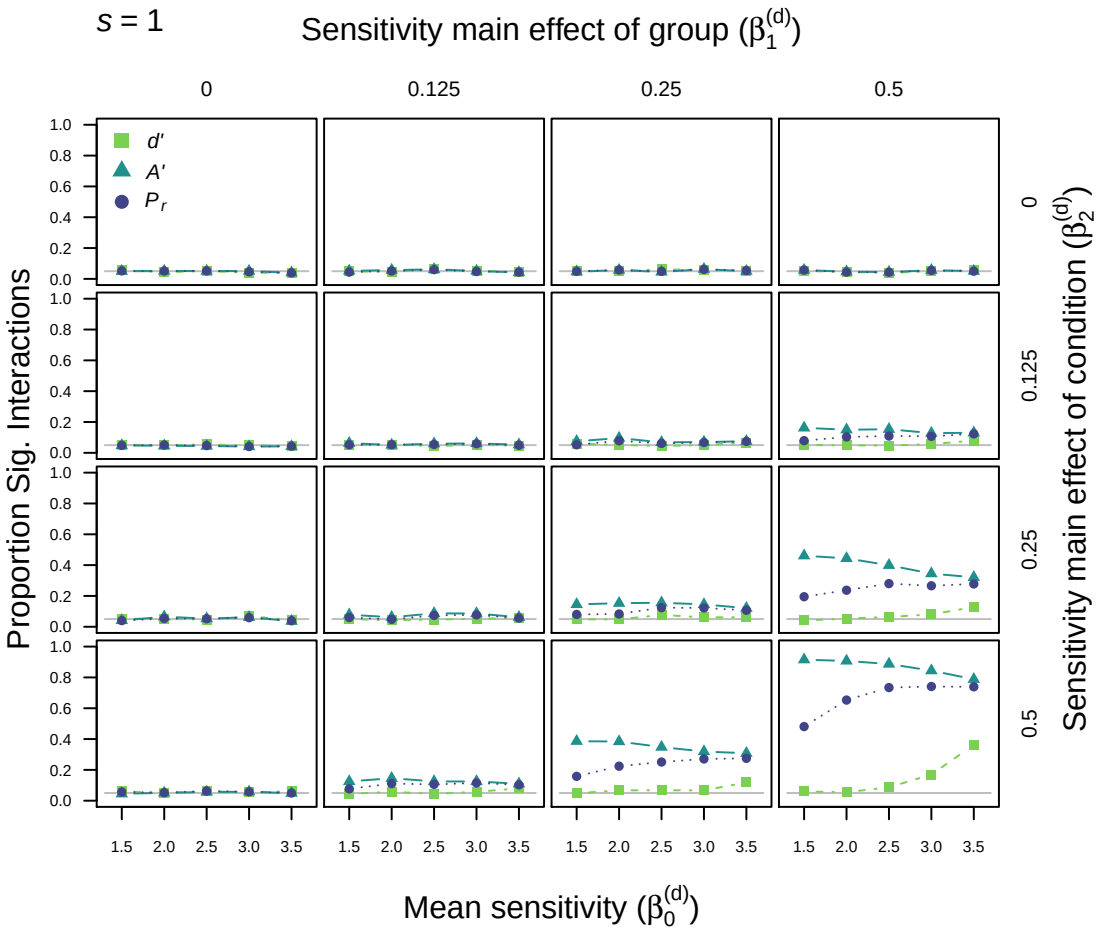

Figure 8: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall detection sensitivity (x axis), magnitude of main effect of group (left to right) and condition (top to bottom) on sensitivity.

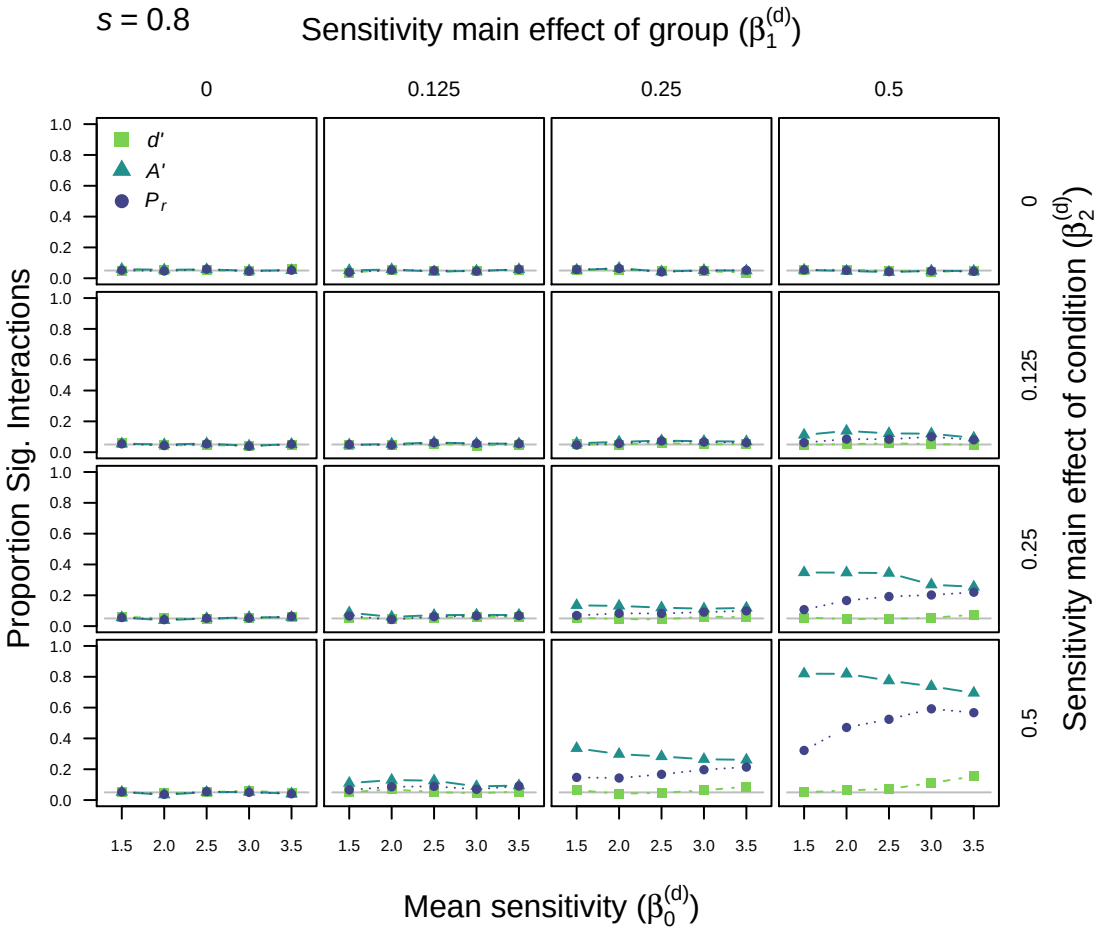

Figure 9: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effect of group (left to right) and condition (top to bottom) on sensitivity.

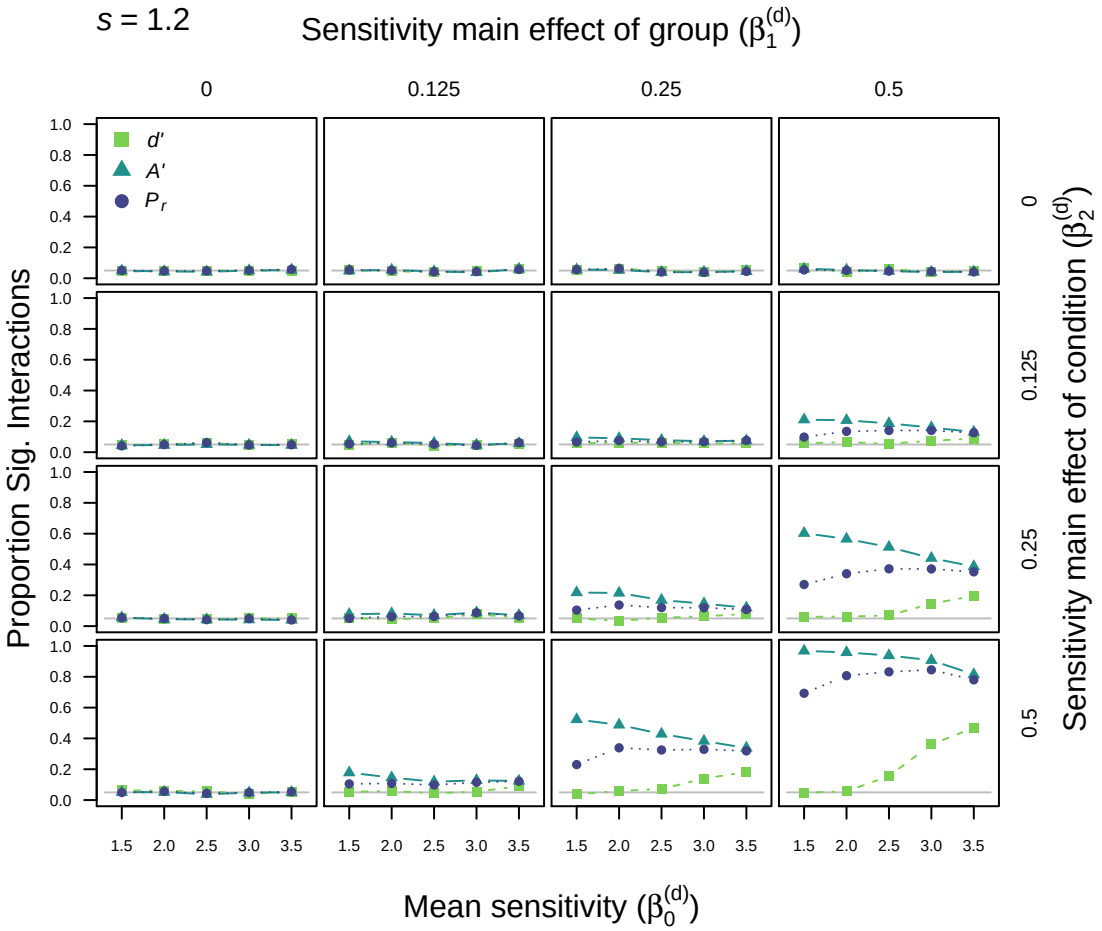

Figure 10: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effect of group (left to right) and condition (top to bottom) on sensitivity.

4 Simulation 4: introduce individual differences in bias and vary grand mean bias

In Simulation 4 we allowed individuals to vary, relative to the grand mean, in their response bias (controlled by the b parameter in our linear equation). The grand mean bias exhibited by observers was also systematically varied as shown in Figures 11 to 14.

4.1 Two-high threshold

Main effects of group and condition on sensitivity (P_r) were fixed to be equal and grand mean bias ($\beta_0^{(B_r)}$) was allowed to vary from conservative to liberal guessing biases (0.3, 0.4, 0.5, 0.6, 0.7. Top to bottom panels of Figure 11). Individual guessing biases were made to vary around the grand mean by sampling from a normal distribution with a mean of 0 and standard deviation ($\sigma_S^{(B_r)}$) of 0.1 (the same variation used for the detection parameter, P_r).

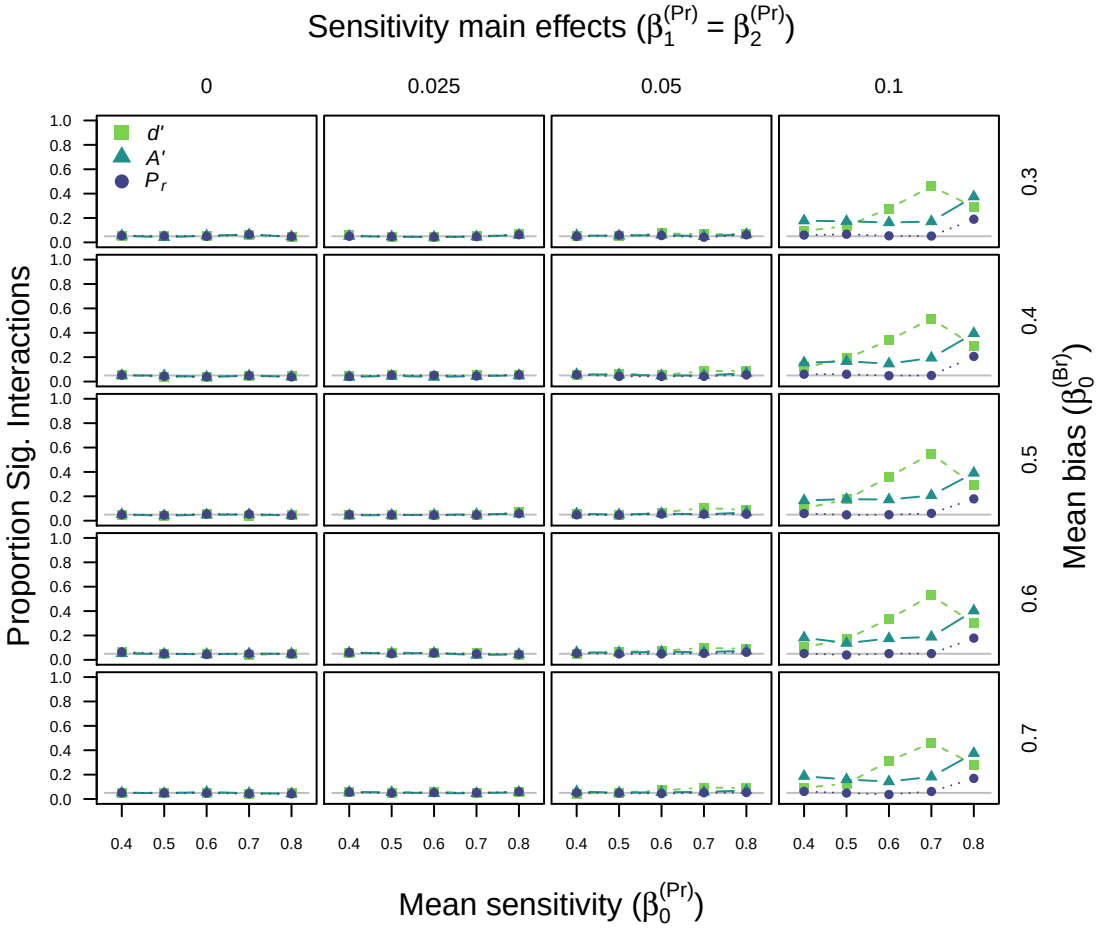

Figure 11: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall detection probability (x axis), magnitude of main effects of group and condition on detection (left to right), and overall bias to responding target versus non-target (top to bottom).

4.2 Signal detection

For the signal detection simulations, overall criterion placement ($\beta_0^{(k)}$) was varied from liberal to conservative (-1, 0.5, 0, 0.5, 1. Top to bottom panels of Figure 12) and individual participant criterion placement deviated from this normally with a standard deviation ($\sigma_S^{(k)}$) of 0.5.

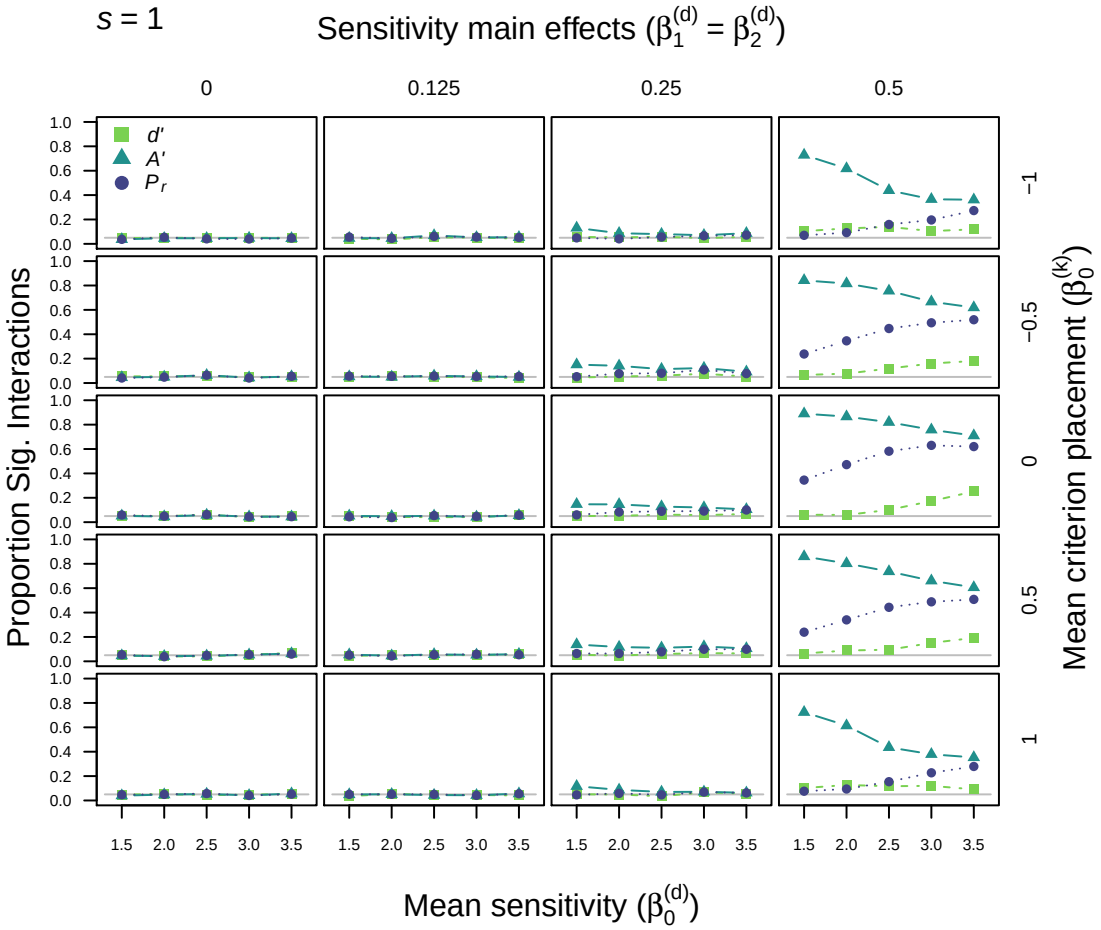

Figure 12: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall detection sensitivity (x axis), magnitude of main effects of group and condition on sensitivity (left to right) and overall criterion placement (top to bottom).

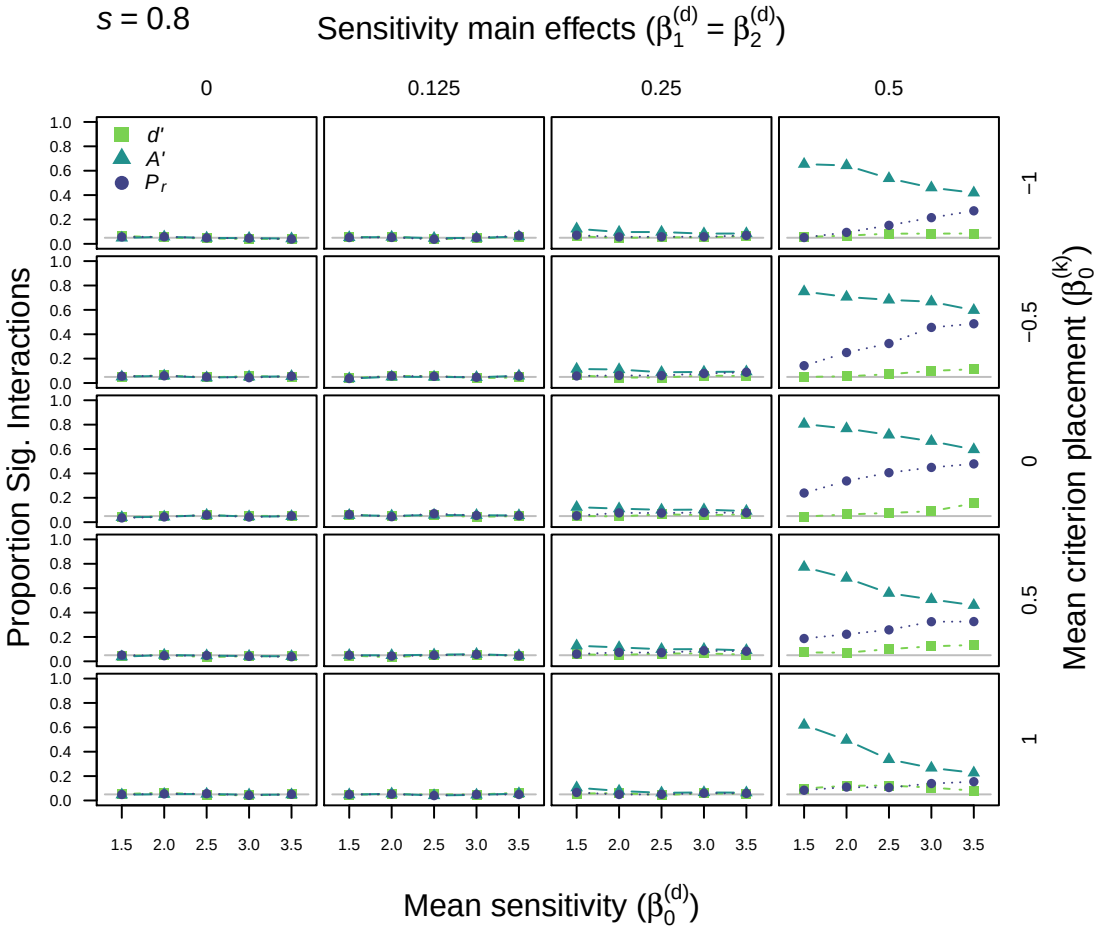

Figure 13: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects of group and condition on sensitivity (left to right) and overall criterion placement (top to bottom).

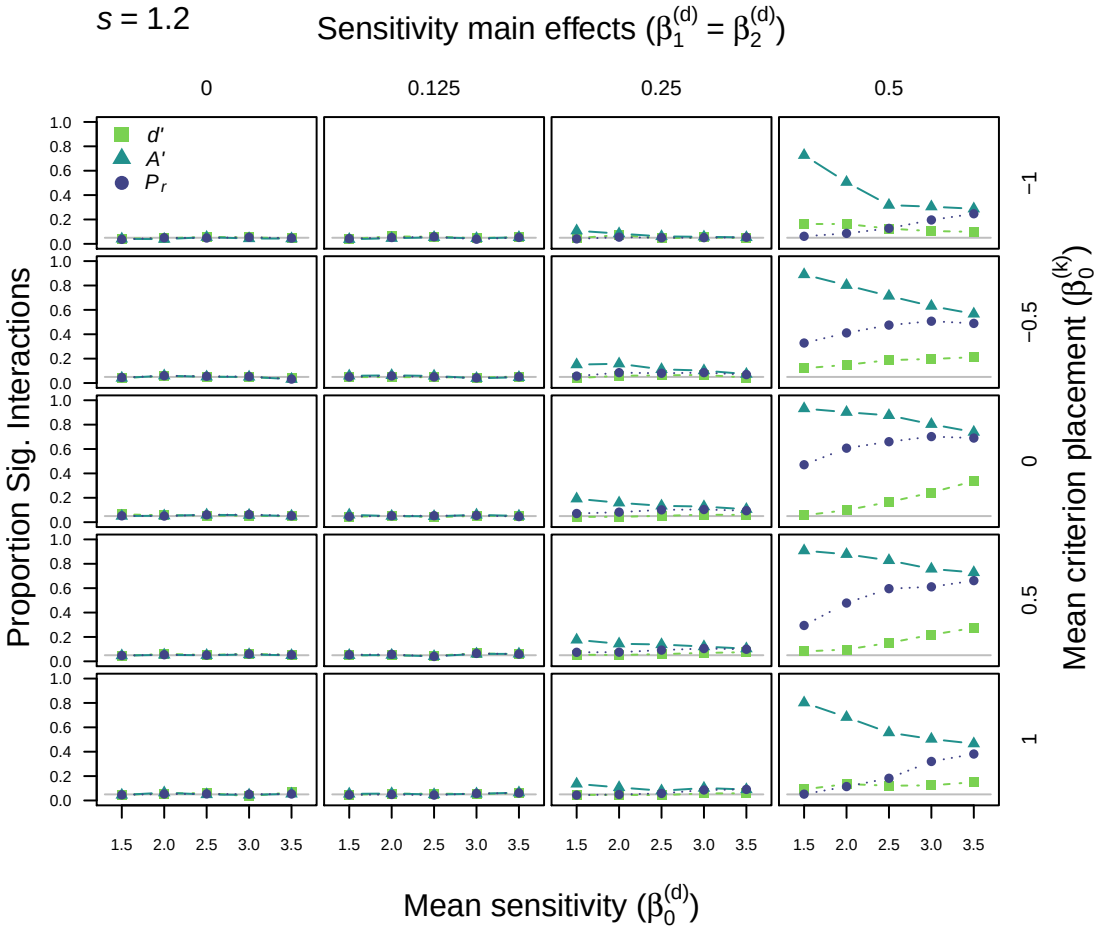

Figure 14: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall detection sensitivity (x axis), magnitude of main effects of group and condition on sensitivity (left to right) and overall criterion placement (top to bottom).

5 Simulation 5: allow group differences in bias

To build on simulation 4, we considered situations where groups differ in their biases towards responding *target* or *non-target*. Figure 15 presents the two-high threshold results and Figures 16 to 18 present error rates for signal detection generative models.

5.1 Two-high threshold

In these simulations we fixed the main effects of group and condition on detection probability (P_r) to be 0.1 and, instead, varied the main effect of group ($\beta_1^{(B_r)}$) on guessing bias (0.025, 0.05, 0.1). As depicted in Figure 15, the effect of varying group bias was quite modest and appears to depend on the overall bias exhibited by the sample. At conservative overall biases (top panels), larger group effects on bias increase error rates for d'. At liberal overall biases (bottom panels), the opposite effect is seen.

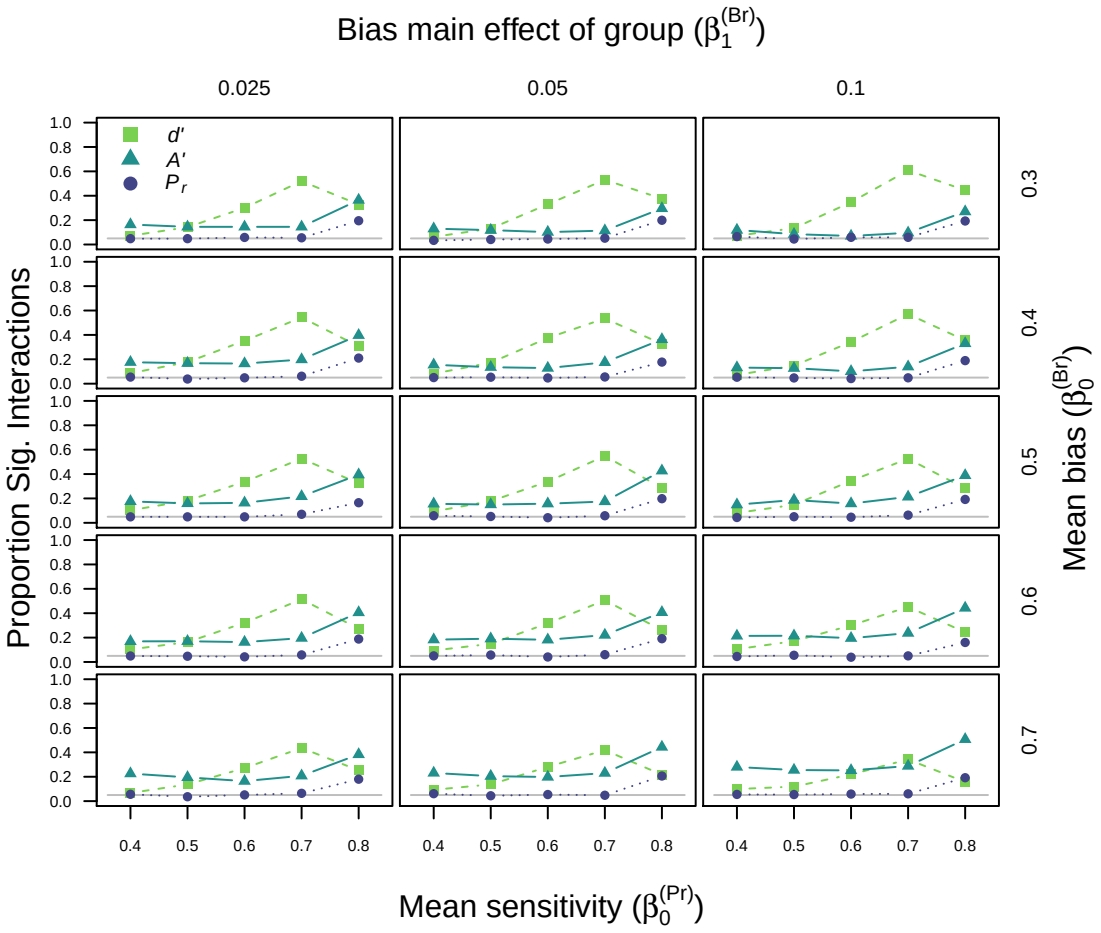

Figure 15: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall detection probability (x axis), magnitude of the main effect of group on guessing bias (left to right), and overall bias to responding target versus non-target (top to bottom).

5.2 Signal detection

For the signal detection generative model, we fixed the two main effects on sensitivity to the medium effect size (0.25) and varied the main effect of group on criterion ($\beta_1^{(k)}$) through 0.125, 0.25, and 0.5. Figures 16 to 18 show the equal and unequal variance simulations. There is little effect of group differences in bias on type I error rate.

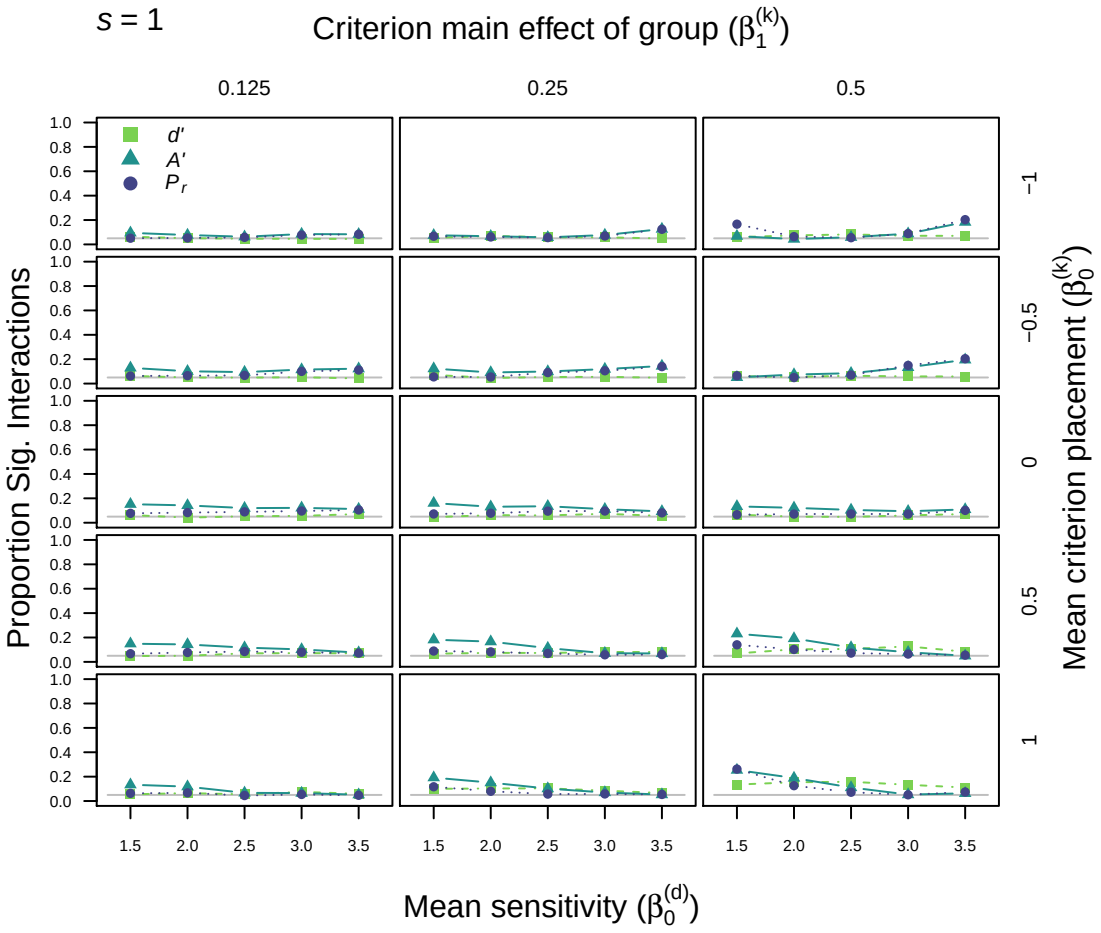

Figure 16: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall detection sensitivity (x axis), magnitude of the main effect of group on criterion (left to right), and overall criterion placement (top to bottom).

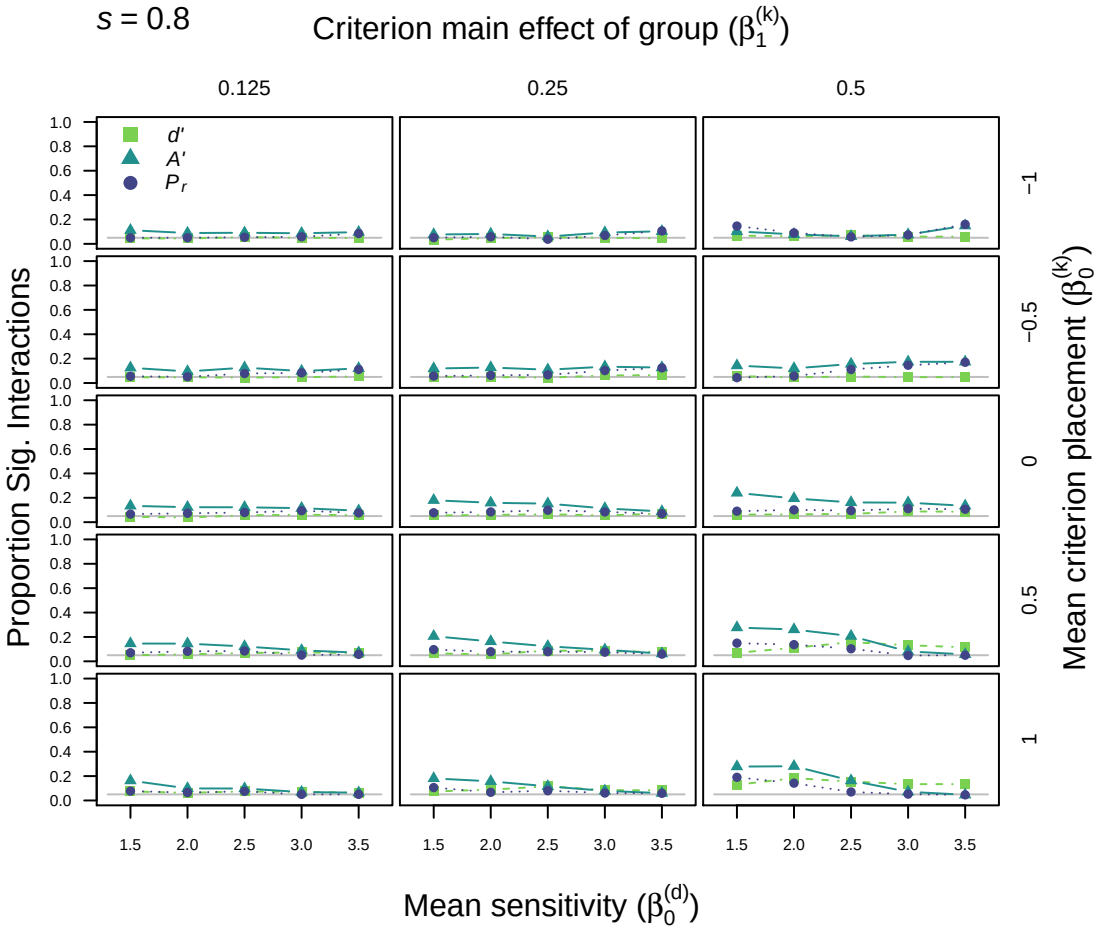

Figure 17: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall detection sensitivity (x axis), magnitude of the main effect of group on criterion (left to right), and overall criterion placement (top to bottom).

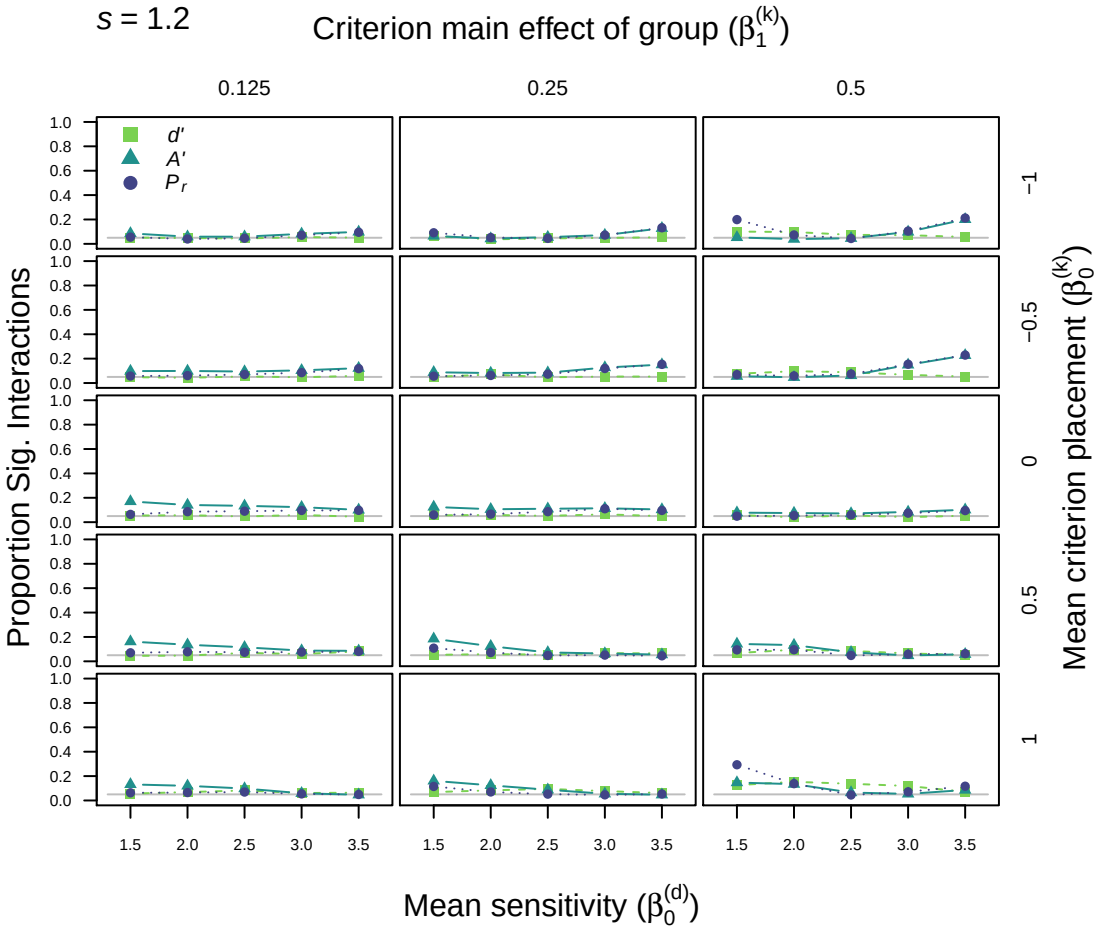

Figure 18: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall detection sensitivity (x axis), magnitude of the main effect of group on criterion (left to right), and overall criterion placement (top to bottom).

6 Simulation 6: orthogonally vary group and condition main effects on bias

These simulations built on Simulation 5 by further introducing main effects of condition, as well as group, on response bias and varying the magnitude of the two main effects orthogonally. Simulation 5 above demonstrated that introducing a main effect of group on response bias had very little influence on the observed error rates. However, the figures below show that allowing for two main effects on bias changes this dramatically. Figure 19 presented these results for a high-threshold model and Figures 20, 21, and 22 present the results for signal detection (equal and unequal variance). We discuss these patterns in detail below.

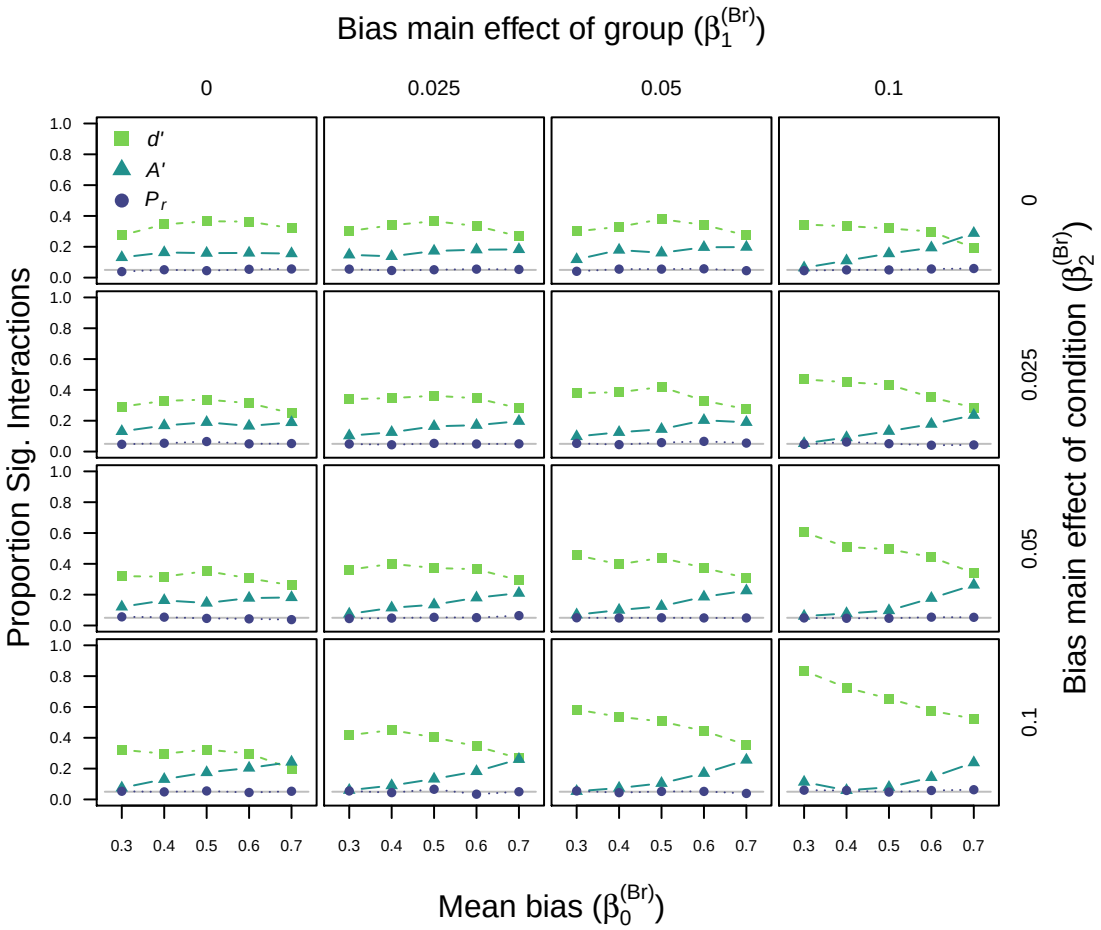

Figure 19: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed).

6.1 Two-high threshold

In this simulation the main effect of group ($\beta_1^{(P_r)}$) and condition ($\beta_2^{(P_r)}$) on the detect probability (sensitivity) was fixed to be 0.1. Grand mean sensitivity ($\beta_0^{(P_r)}$) was also fixed, to 0.6. Rather we varied the grand mean bias ($\beta_0^{(B_r)}$) and orthogonality paired different magnitudes of group ($\beta_1^{(B_r)}$) and condition ($\beta_2^{(B_r)}$) main effects on bias, including no effect (0, 0.025, 0.05, 0.1). The results of this simulation study are presented in Figure 19.

Allowing groups and conditions to differ in guessing bias increases the type I error for interactions when using d', due to the curvature of the ROCs predicted by this measure, but decreases the error associated with using A' (see Figure 19). As the size of the underlying main effects on bias increase (moving from the top left to bottom right of Figure 19), error rates for both measures are increasingly dependent

on the average probability of guessing *target* ($\beta_0^{(B_r)}$). For A', error rates increase as the bias towards responding *target* increases, whereas for d' errors are more likely when observers are biased towards saying *non-target*.

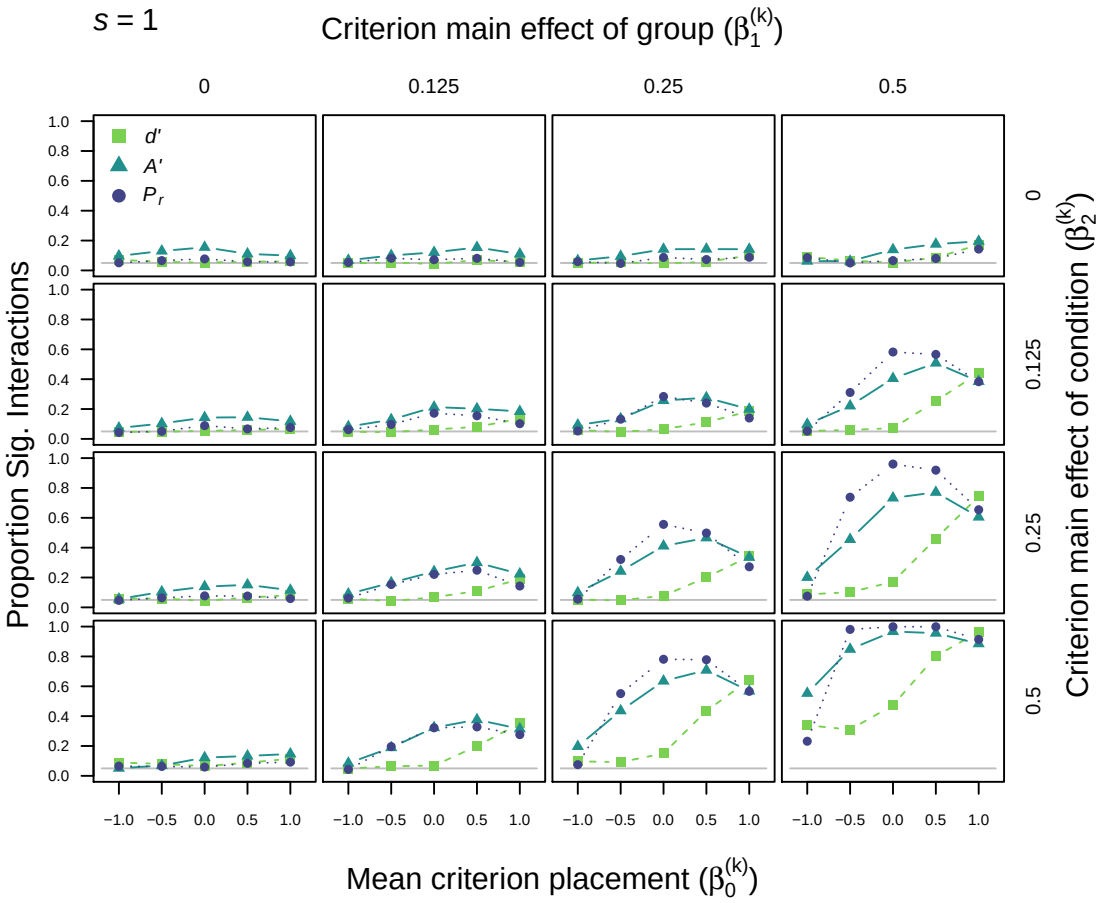

Figure 20: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed).

6.2 Signal detection

For the signal detection simulations we also fixed the magnitude of main effects ($\beta_1^{(d)} = \beta_2^{(d)} = 0.25$) and grand mean sensitivity ($\beta_0^{(d)} = 2$), whilst varying bias. Mean criterion placement ($\beta_0^{(k)}$) was varied from liberal (-1, -0.5) through to conservative (0.5, 1) and main effects of group ($\beta_1^{(k)}$) and condition ($\beta_2^{(k)}$) on criterion placement were also varied (0, 0.125, 0.25, 0.5). Figure 20 presents the resulting estimated type I error rates for tests of interactions on d', A', and P_r for the equal variance generative model. Figures 21 and 22 depict the unequal variance simulations. Relative to the high threshold simulations above, the pattern of error rates is much more complicated.

Focusing on the panels along the top and left sides of Figure 20 shows the situation where there is a *single* main effect on response bias. In this case the only appreciable effect is on A', in that at more extreme main effect sizes the type I

error rate becomes increasingly dependent on the overall level of bias exhibited by observers (along the x-axis). However, error rates do not exceed 0.2.

The remaining 9 panels depict the case where both group and condition have an effect on response bias. An interesting pattern emerges, in that the error rate for A' and especially for P_r becomes increasingly like an inverted 'U' in shape, with more extreme errors when the average criterion ($\beta_0^{(k)}$) is neutral. Type I error rates drop to 0.05 as criterion becomes more liberal and do not show an equivalent drop for more conservative responding. d' also shows unacceptable error rates with large bias effects and a general trend towards conservative responding.

The cause of the shape of error rates as a function of mean response bias can be briefly summarized as follows. The inverted 'U' shape arises due to the fact that the curvature of the Gaussian signal detection ROC is largest when overall responding is neutral, making the difference between groups and conditions clearer and, hence, the disagreement between measures more pronounced. The asymmetry arises due to our parameter settings; using positive values for both sensitivity main effects and bias main effects results in more sensitive groups and conditions exhibiting more conservative responding. This results in a pronounced floor effect—as the most sensitive conditions will produce few if any false-alarms—causing errors for all measures, including d'. As we discuss in the main manuscript, having individual rates of 1 or 0 (even when overall performance is off ceiling or floor) can greatly distort conclusions regarding interactions, even when an appropriate measure is used.

To summarize, adding variation in bias between groups and conditions further complicates the interpretation of interactions on measures of sensitivity. For a threshold generative model, error rates for d' are particularly pronounced (greater than 0.8) when overall participants are conservative and there are large main effects on bias. For A' errors become increasingly prevalent at liberal guessing biases as the size of main effect on response bias increases. When the underlying model is equal variance Gaussian signal detection, error rates for A' and P_r are maximized at liberal overall criterion placements where the ROC curvature is greatest and hence disagreement between measures is greatest. As shown below, simulations with unequal variance versions of the signal detection model revealed the same general inverted U shaped type I error functions for the incorrect measure. As is to be expected, the precise shape of this function depends on the direction of the unequal variance; more variable targets ($s = 0.8$. Figure 21) shift the function somewhat to the right, whereas more variable non-targets ($s = 1.2$. Figure 22) shift the function to the left, producing somewhat more extreme error rates for all three measures

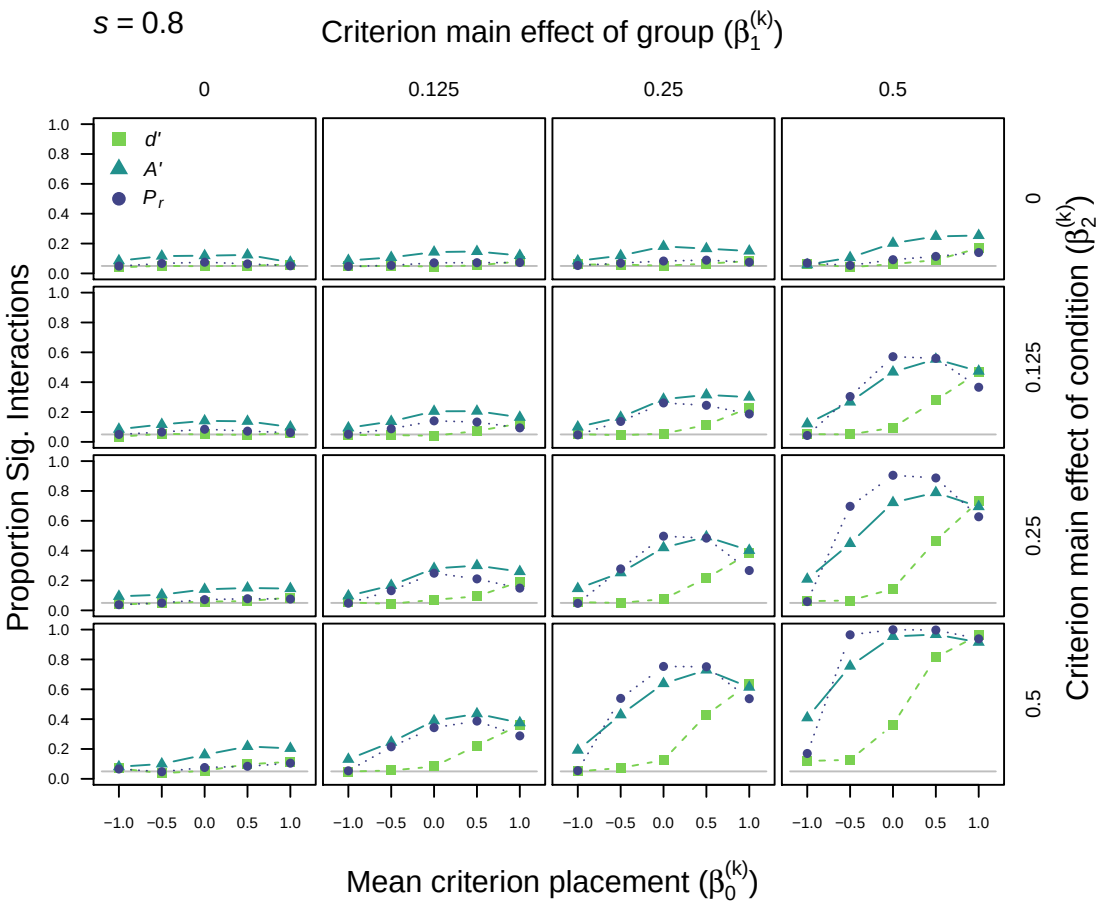

Figure 21: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed).

relative to the equal variance case.

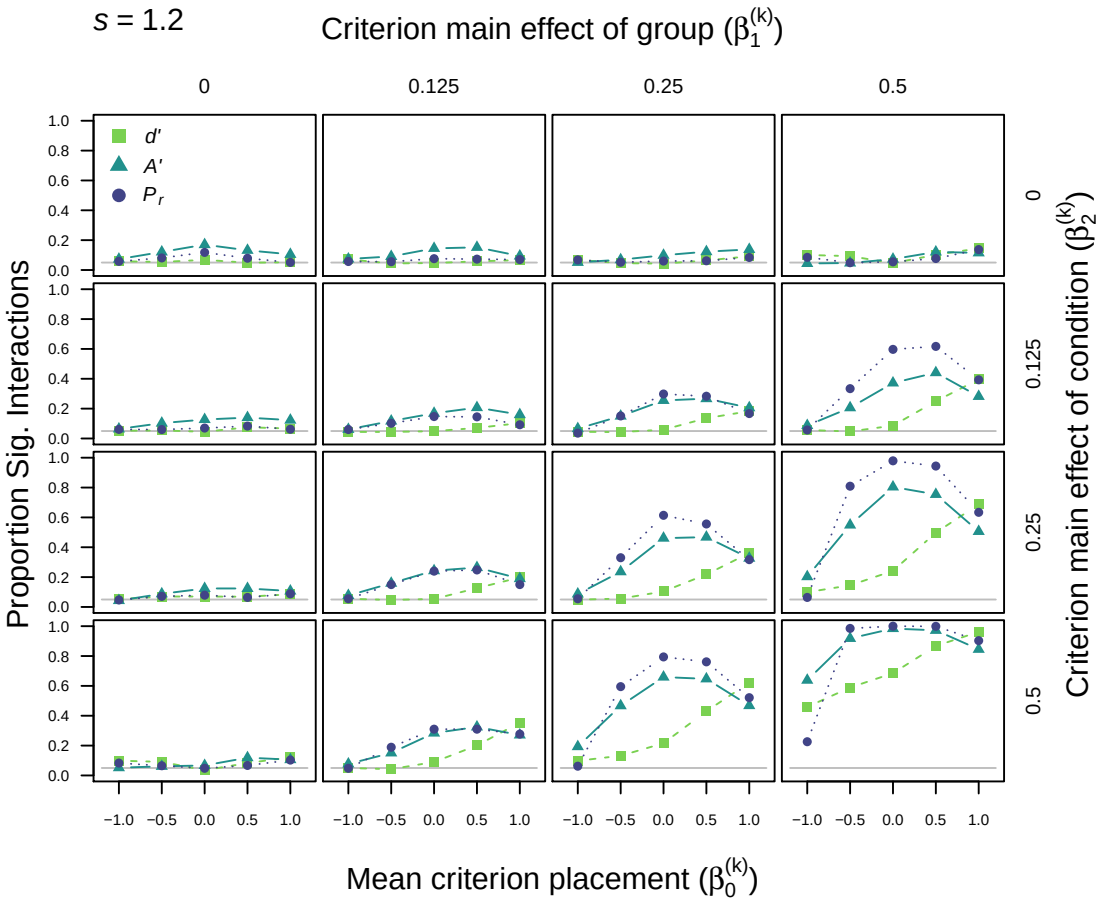

Figure 22: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed).

7 Simulation 7: vary group and condition main effects on bias with no sensitivity main effects

Simulation 7 was a logical extension of simulation 6 and previous assessments of type I error for main effects (Schooler & Shiffrin, 2005; Rotello et al., 2008). These simulations were identical to those reported in Simulation 6 above, with the exception that in this case we removed the main effects of group and condition on *sensitivity*. Figures 23 to 26 show the results of this. It is clear that unacceptable type I error rates can also arise in the *absence* of main effects on sensitivity, provided that there are differences in response bias, complementing the earlier findings for main effects.

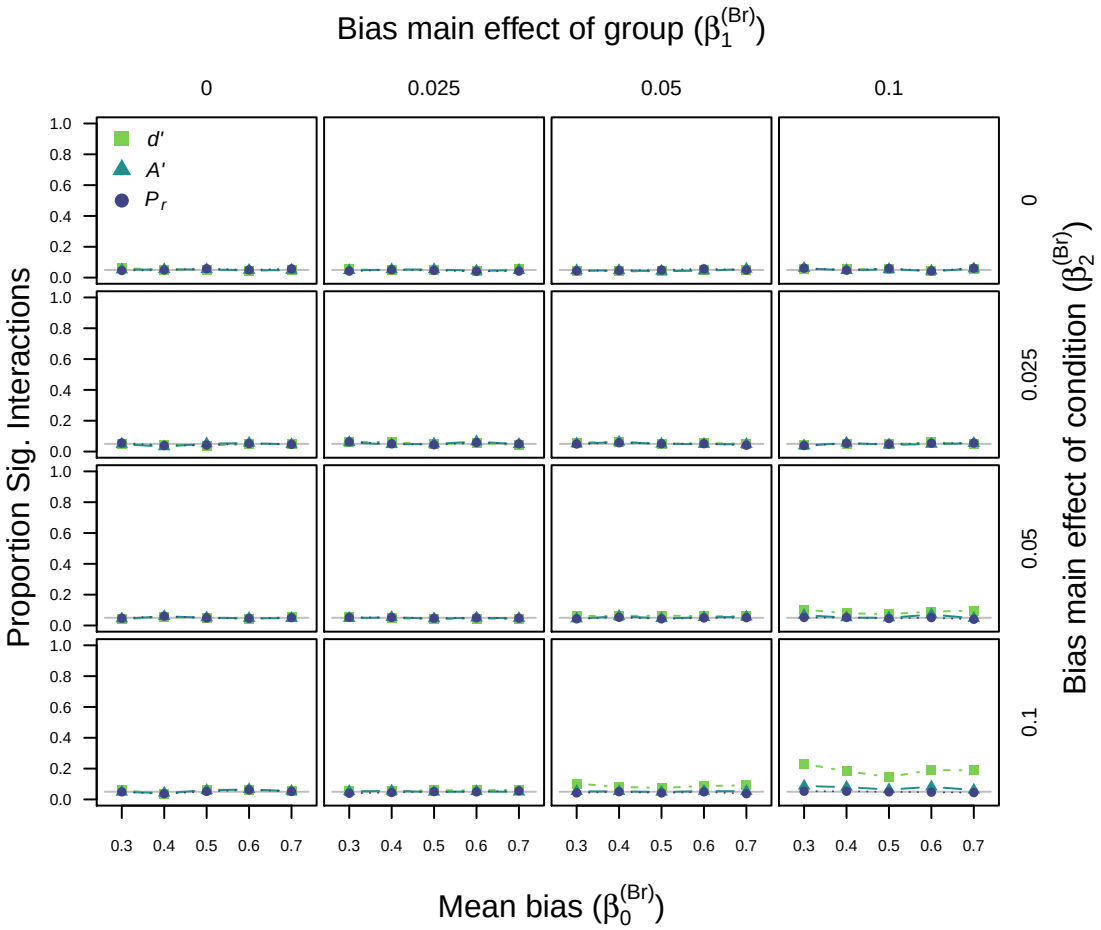

Figure 23: Type I error rates for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed). Note: in this case there were no underlying main effects on P_r.

7.1 Two-high threshold

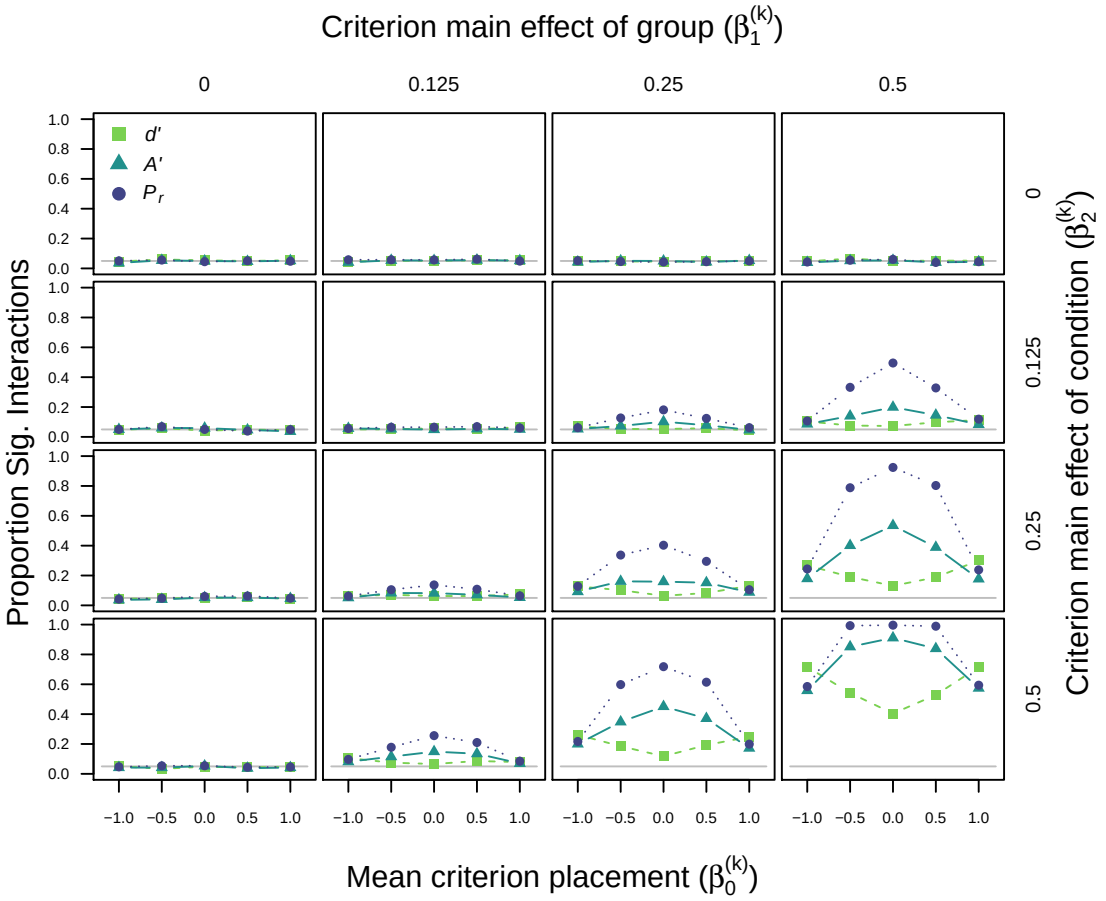

Figure 24: Type I error rates for d', A', and P_r with an underlying Gaussian equal-variance signal detection theory (EV-SDT) model. This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed). Note: in this case there were no underlying main effects on d.

7.2 Signal detection

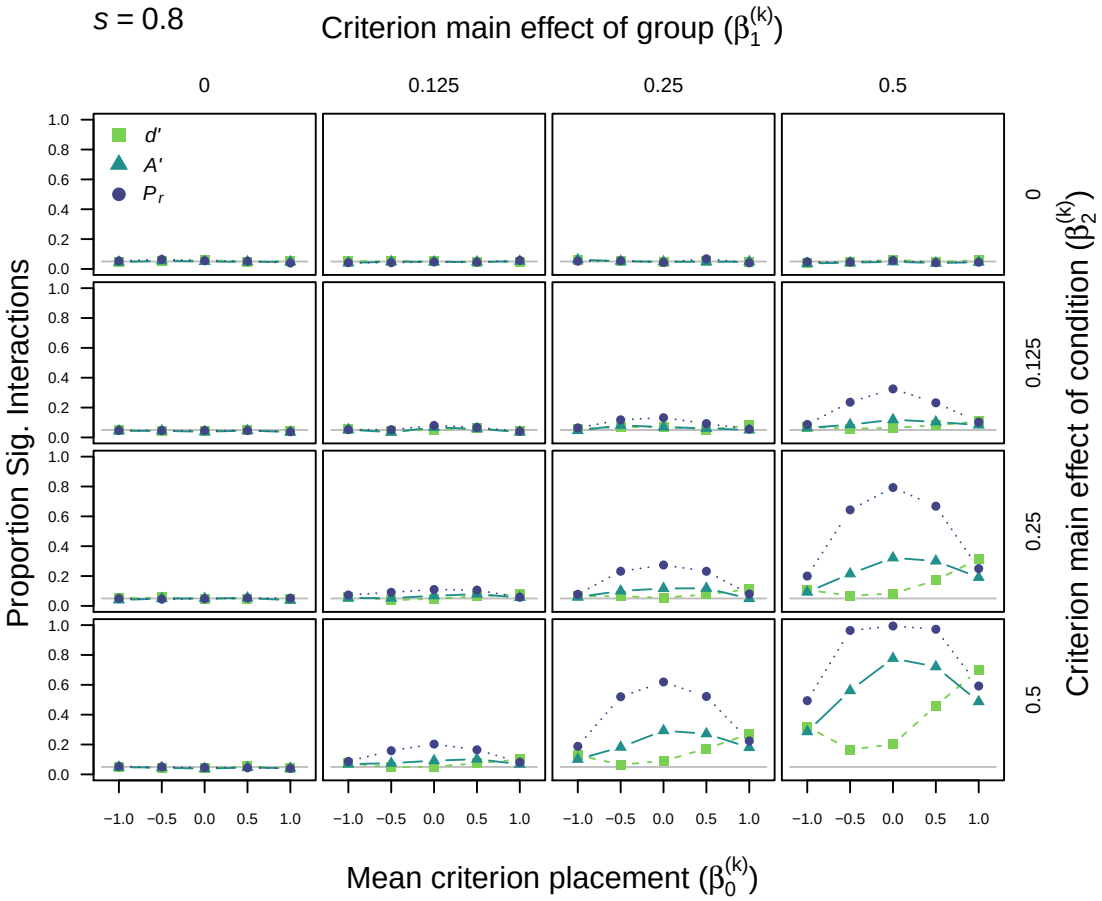

Figure 25: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed). Note: in this case there were no underlying main effects on d.

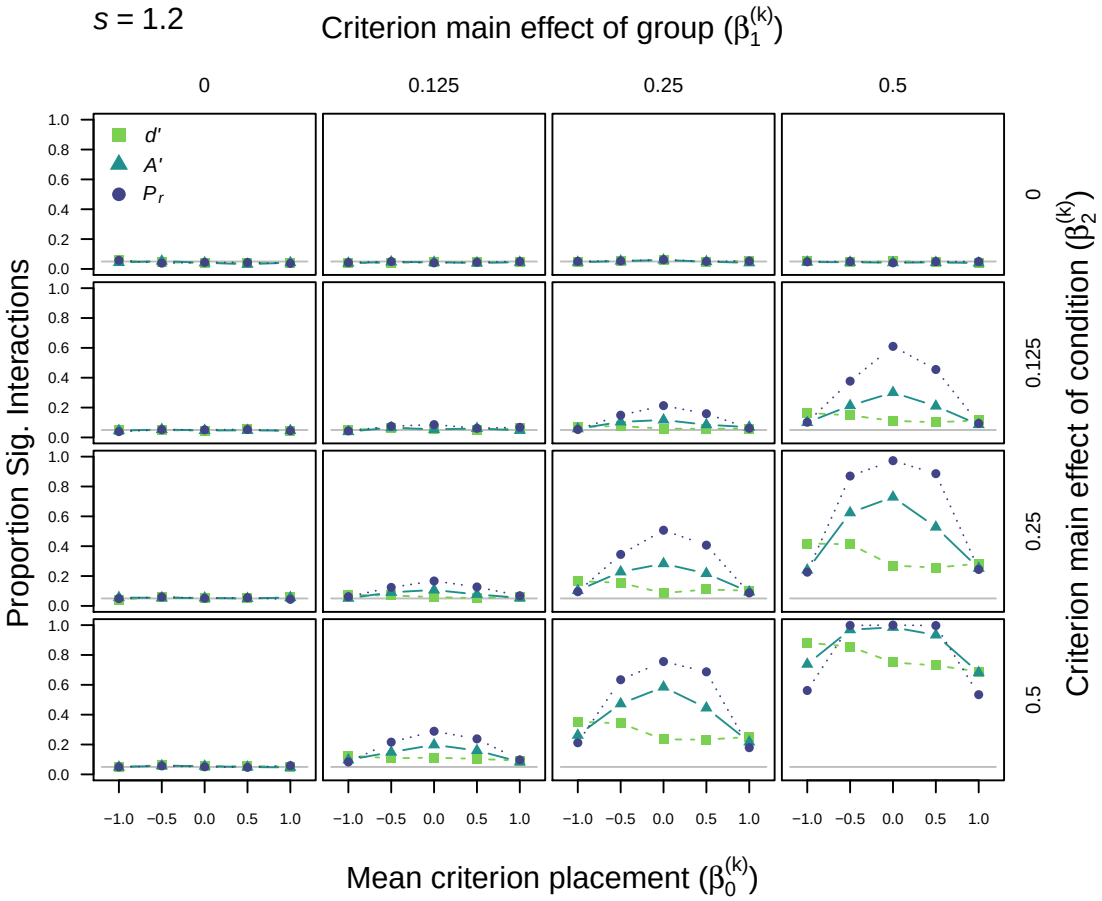

Figure 26: Type I error rates for d', A', and P_r with an underlying Gaussian unequal-variance signal detection theory (UEV-SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall bias (x axis) and the main effects of group (left to right) and condition (top to bottom) on bias (sensitivity grand mean and main effects were fixed). Note: in this case there were no underlying main effects on d.

Power Simulations

8 Simulation 8: varying overall sensitivity, main effects, and number of participants

In simulations 8–12 we assessed the power of the three sensitivity measures under different generative models. To begin with, in simulation 8, we assessed the influence of the magnitude of main effects on sensitivity and the number of subjects on power. Here there is an additional variable, namely the magnitude of the true interaction effect, so the power simulations were much larger than the type I simulations above. As such in the manuscript we only presented results from simulations in which the main effects on sensitivity was largest (and hence disagreement between measures is the most pronounced). Here we present plots from all main effect magnitudes considered. Figures 27 to 36 present results from the 8th set of simulations that were not presented in the main manuscript.

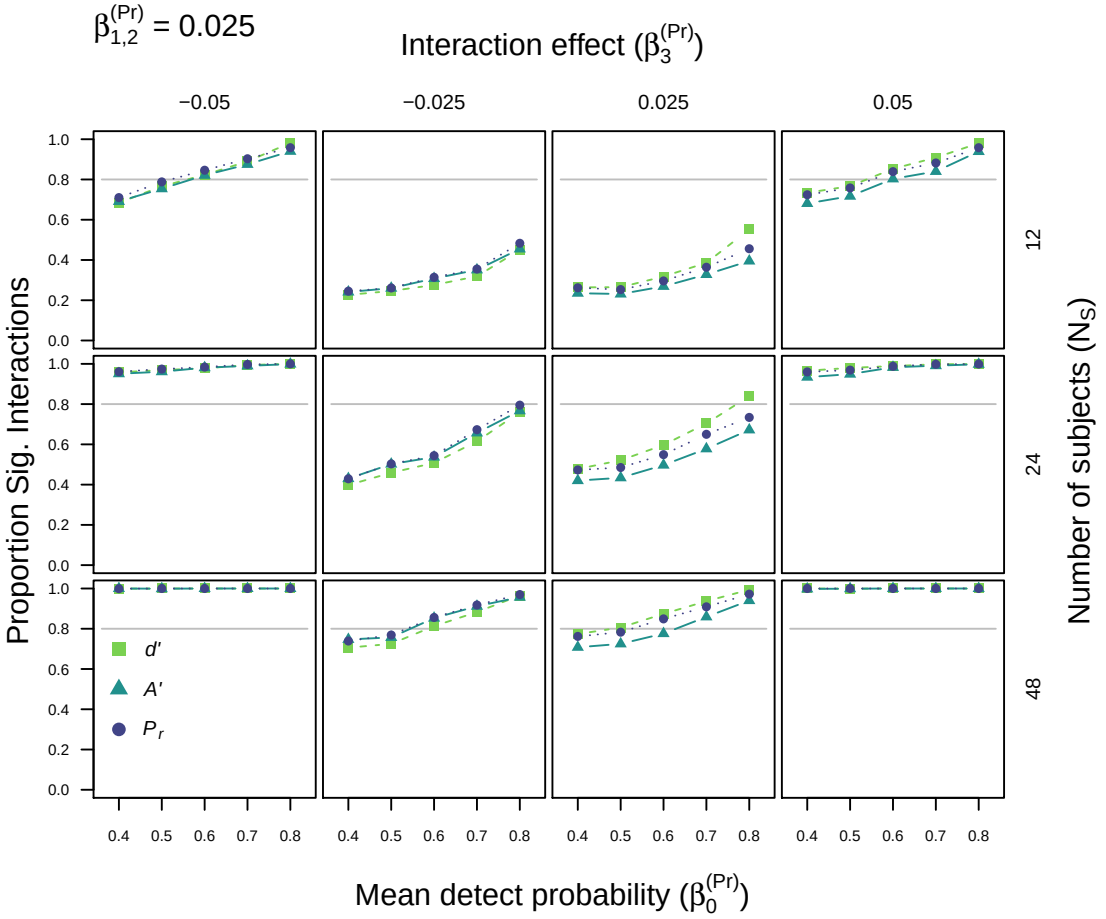

Figure 27: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.025.

8.1 Two-high threshold

For these simulations we returned to the initial parameter settings used in the first type I error simulations, with the additional variation in the interaction parameter, $\beta_3^{(P_r)}$. Figures 27 and 28 depict the two-high threshold power simulations with small and medium main effects on sensitivity (P_r), respectively. The results with the largest main effects are reported in the main manuscript. Note that in these simulations bias was not varied between individuals, groups, or conditions ($B_r = 0.5$).

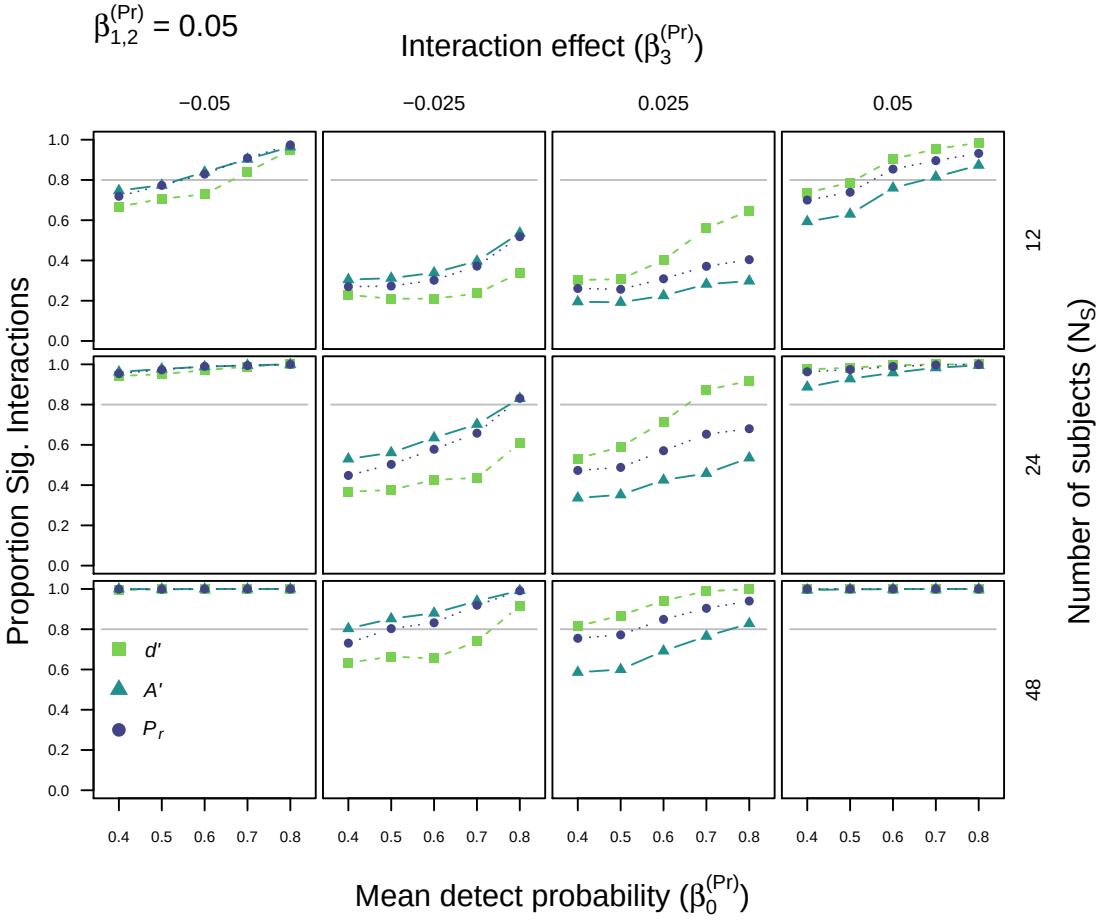

Figure 28: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.05.

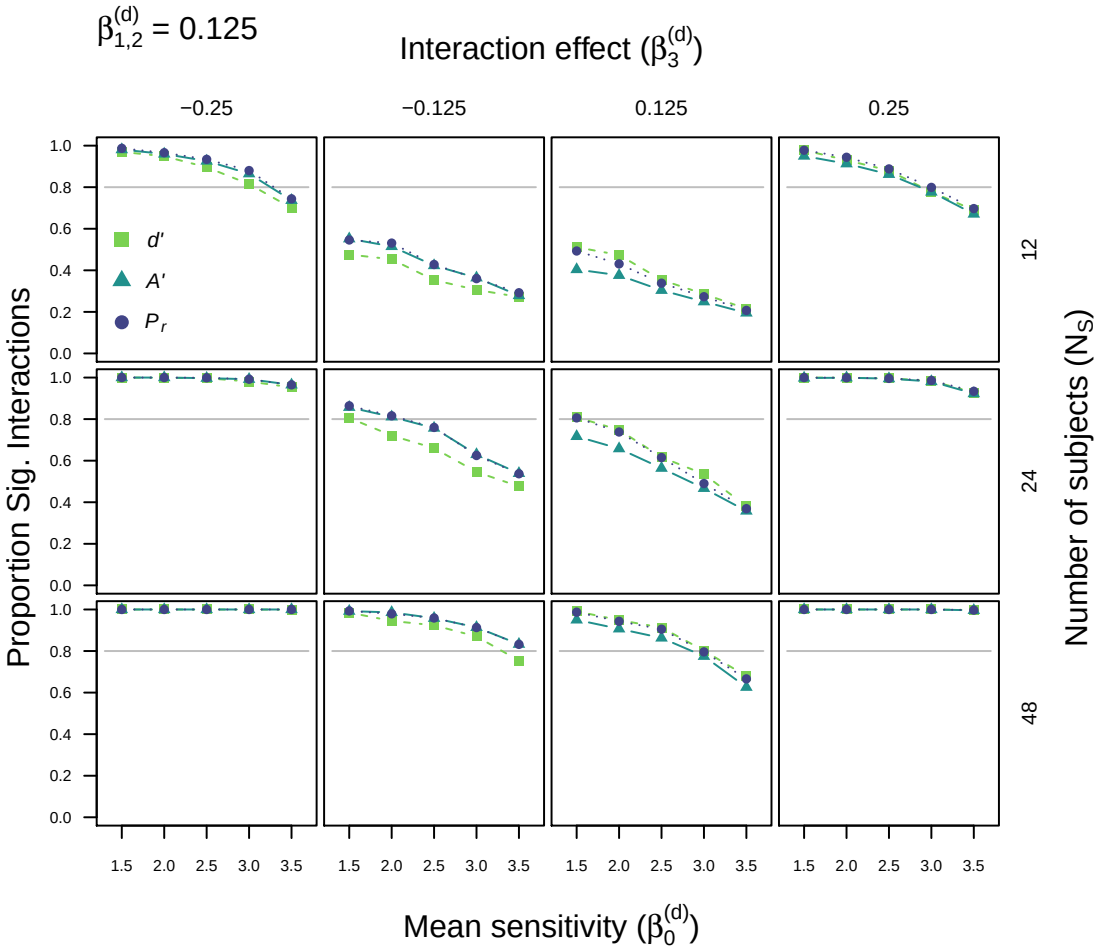

Figure 29: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

8.2 Signal detection

Figures 29 and 29 depict the equal variance signal detection simulations with small and medium main effects on sensitivity (d), respectively. The largest main effect size is reported in the main manuscript. Criterion was not varied between individuals, groups, or conditions ($k = 0$) in these simulations.

Figures 31 to 36 present the unequal variance simulations for the three sensitivity effects sizes. Parameter settings are given on each of these figures.

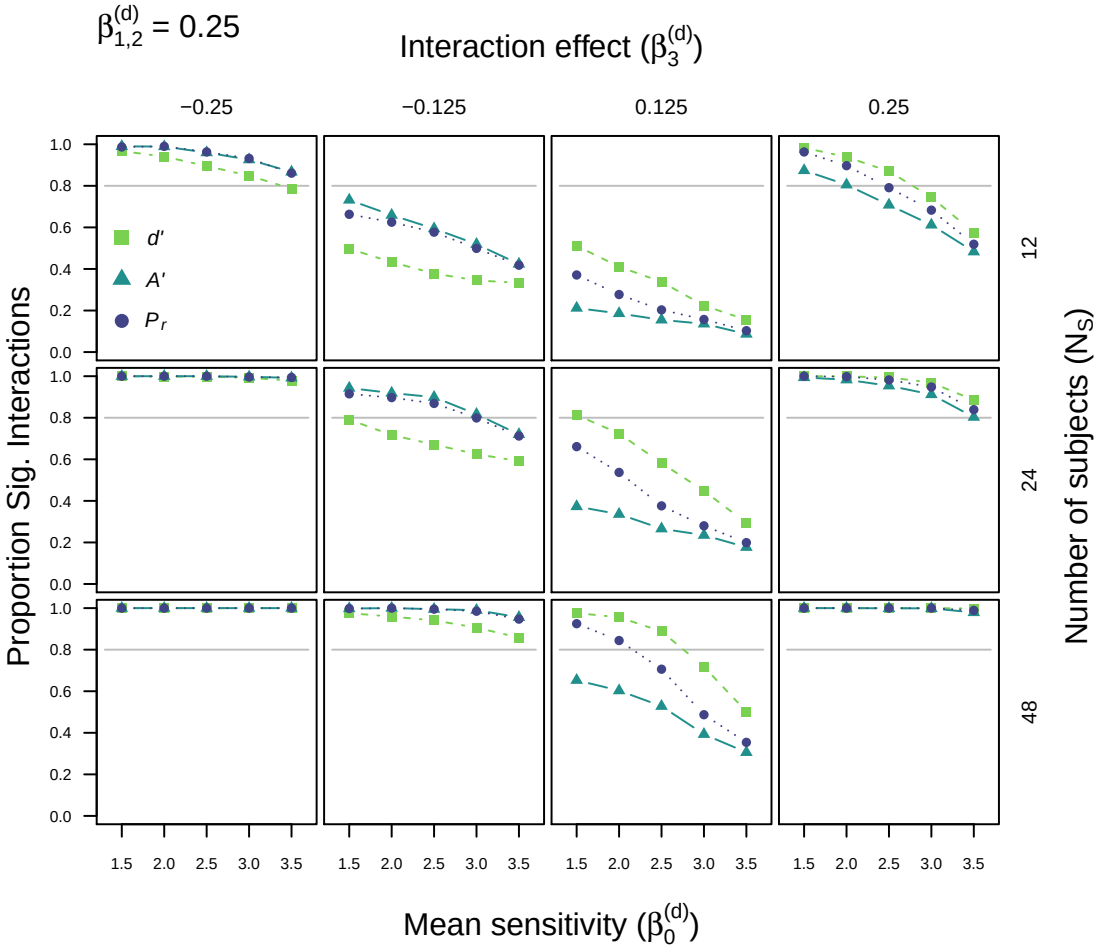

Figure 30: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

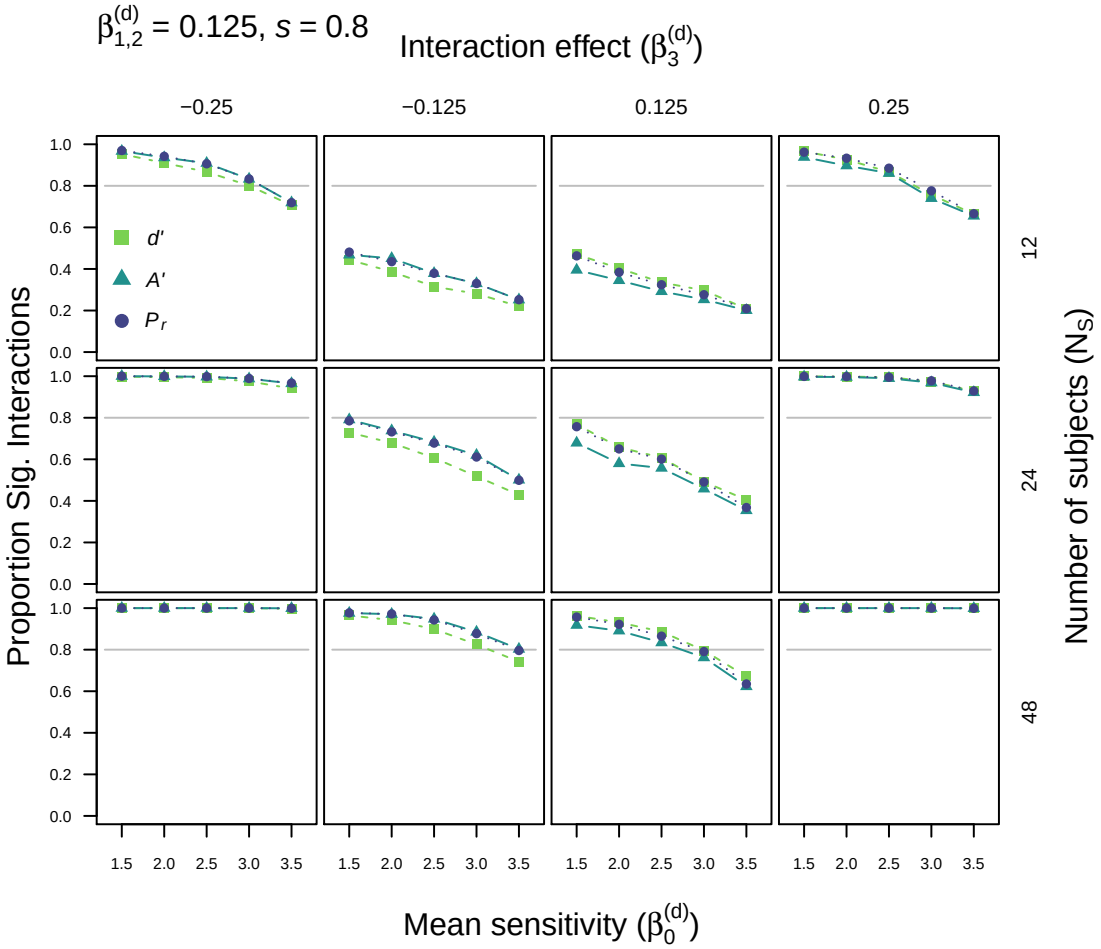

Figure 31: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

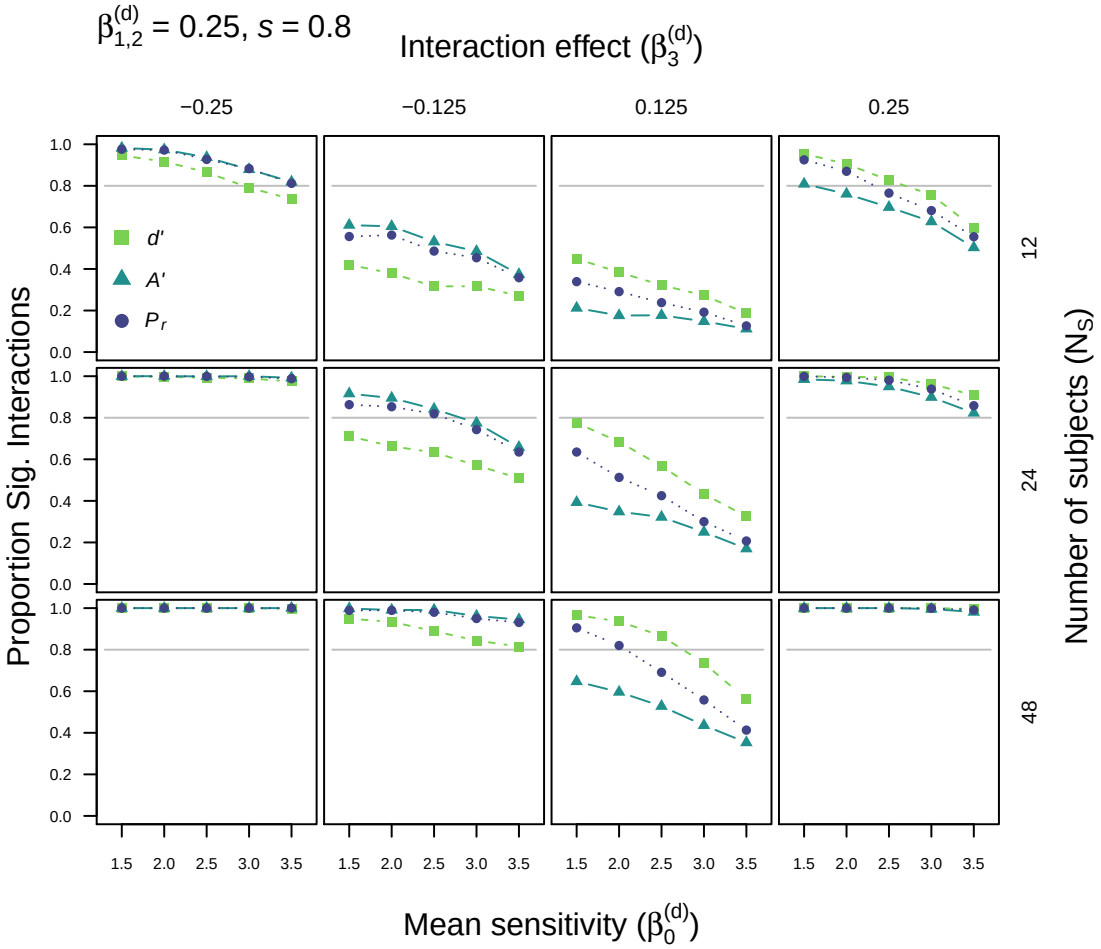

Figure 32: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

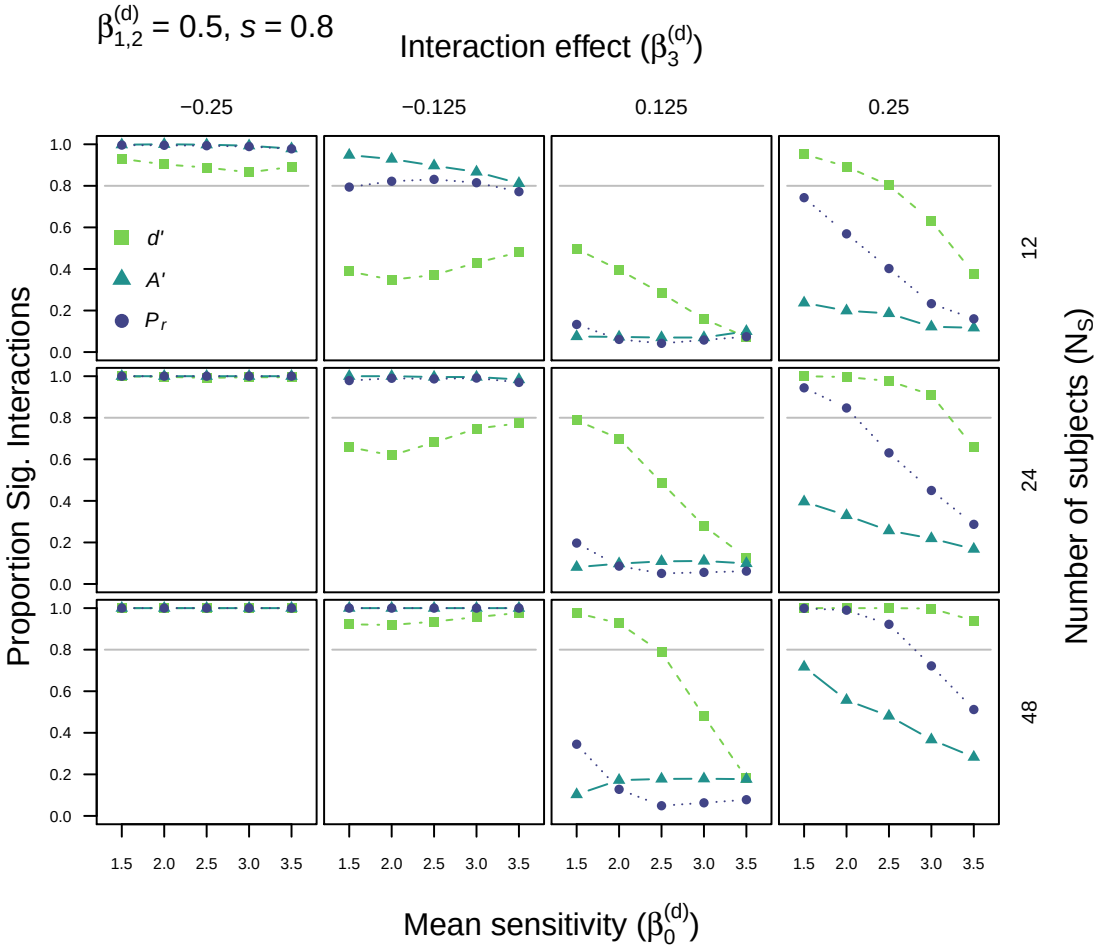

Figure 33: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

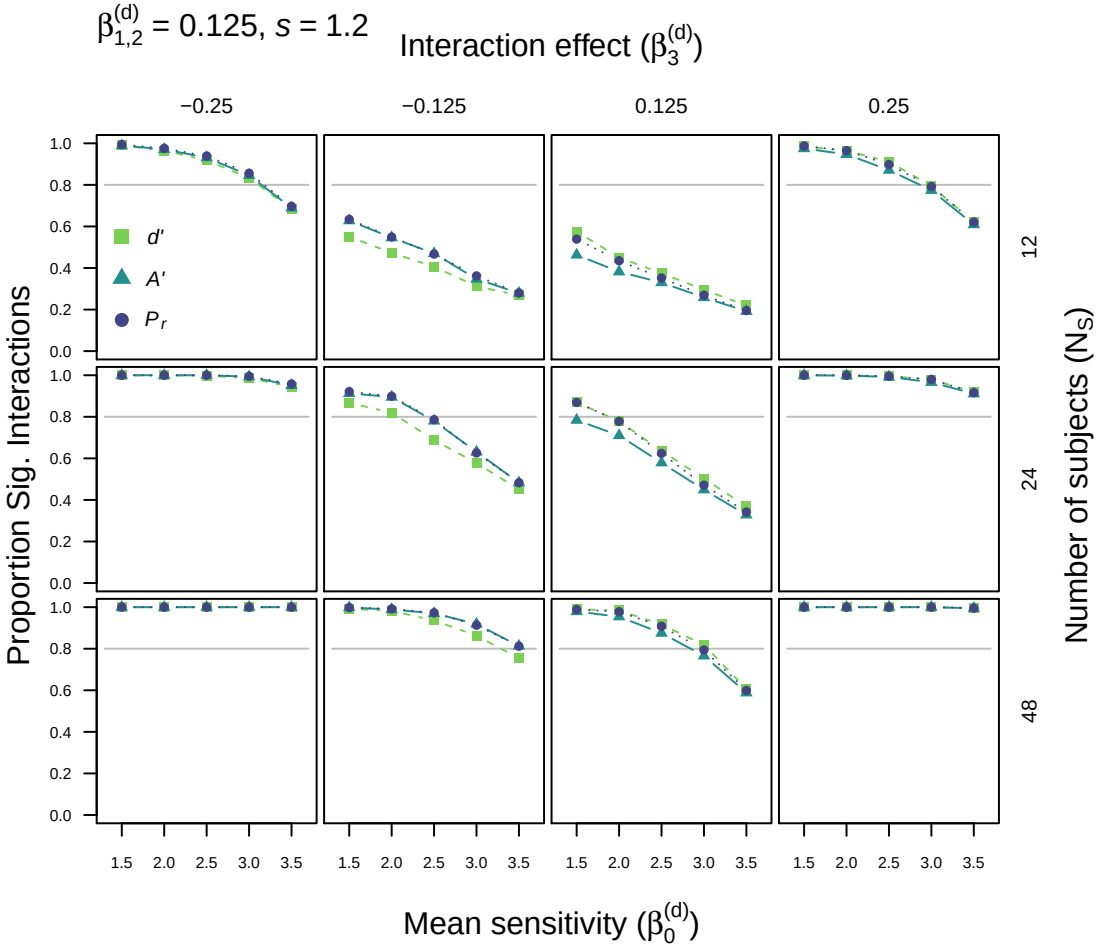

Figure 34: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

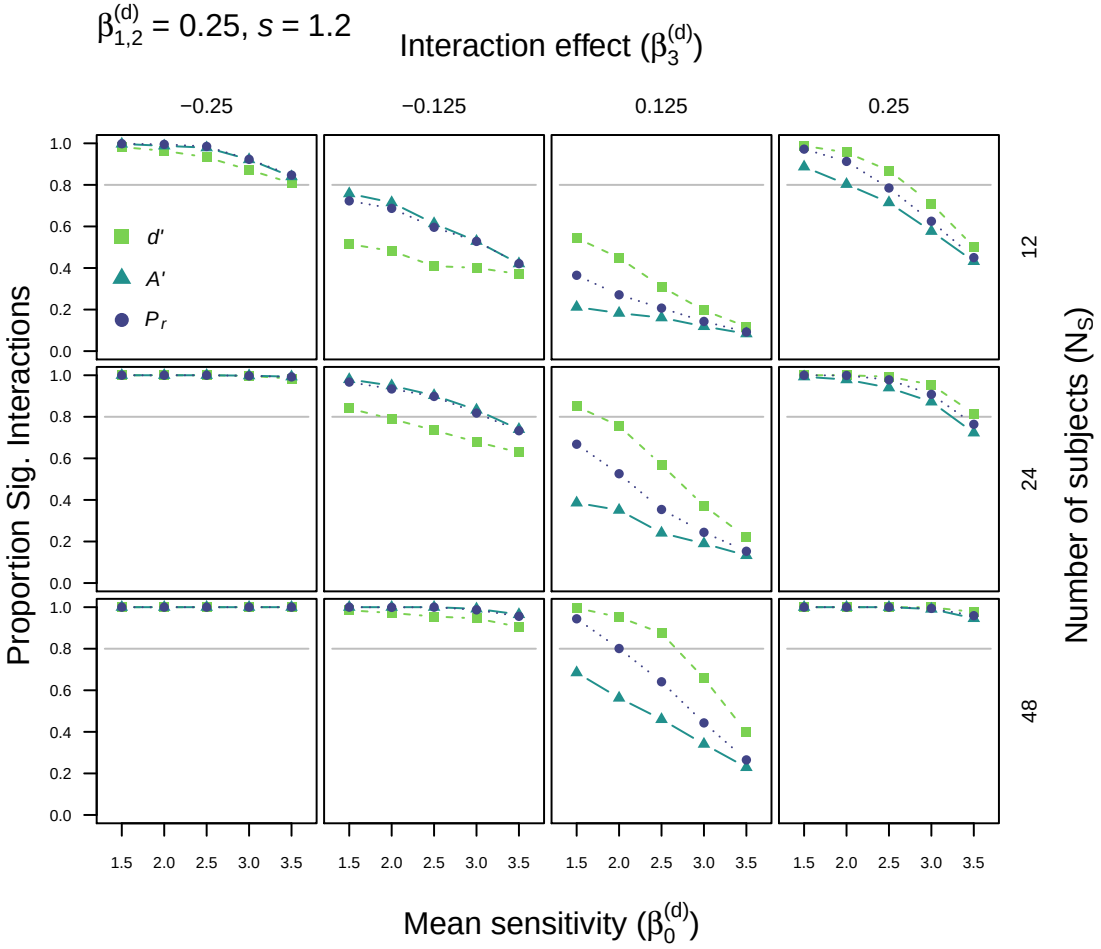

Figure 35: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

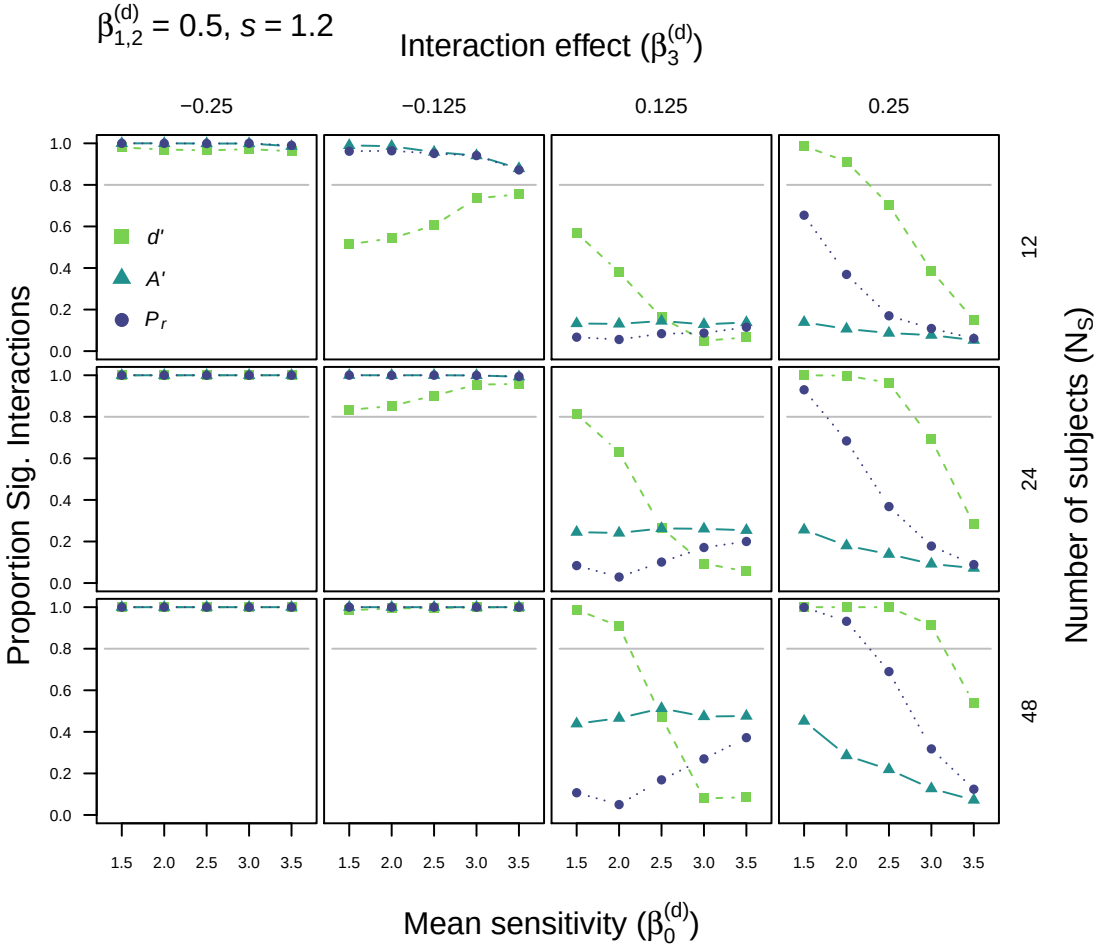

Figure 36: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of participants per group (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

9 Simulation 9: varying overall sensitivity, main effects, and number of trials

Rather than vary the number of participants per group, simulation 9 instead fixed this to 24 and examined power with various numbers of trials per condition (12, 24, 48). Otherwise the remaining parameter settings were the same as simulation 8. Figures 37 to 48 present the results of this set of simulations.

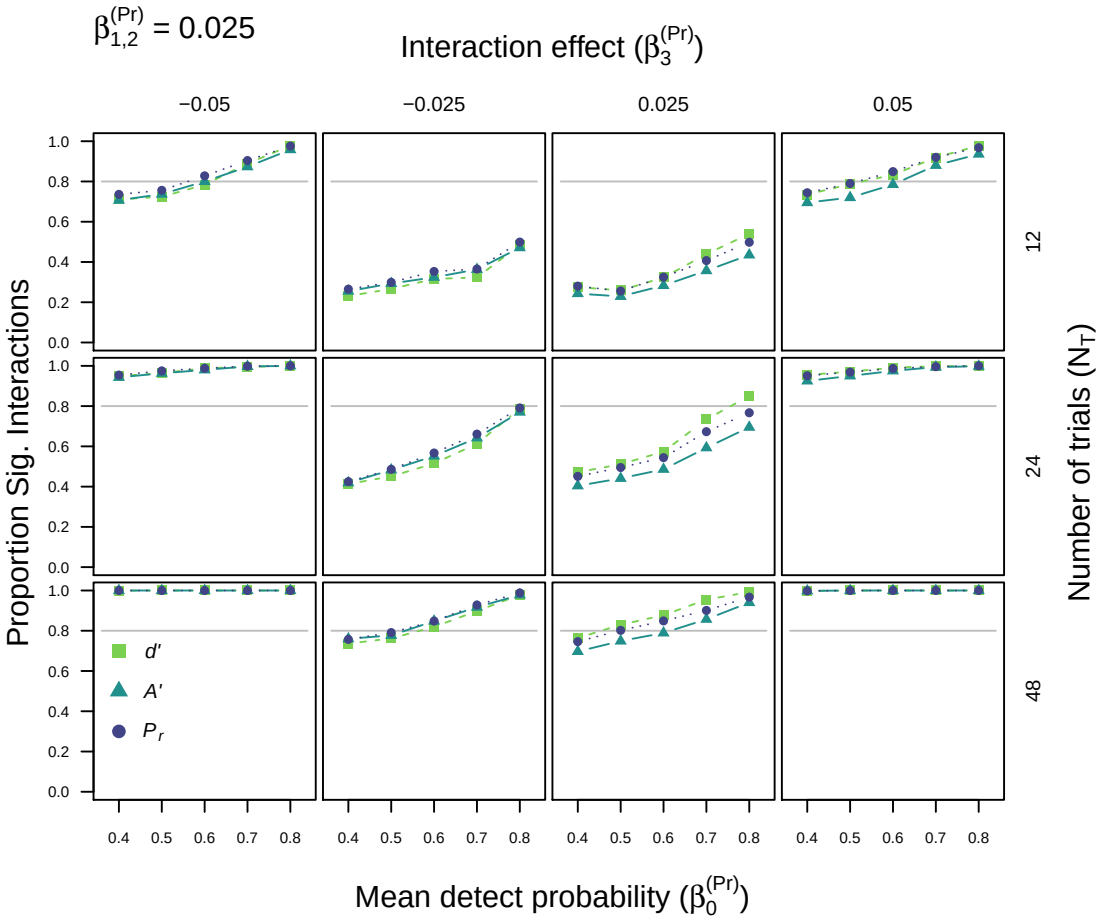

Figure 37: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.025.

9.1 Two-high threshold

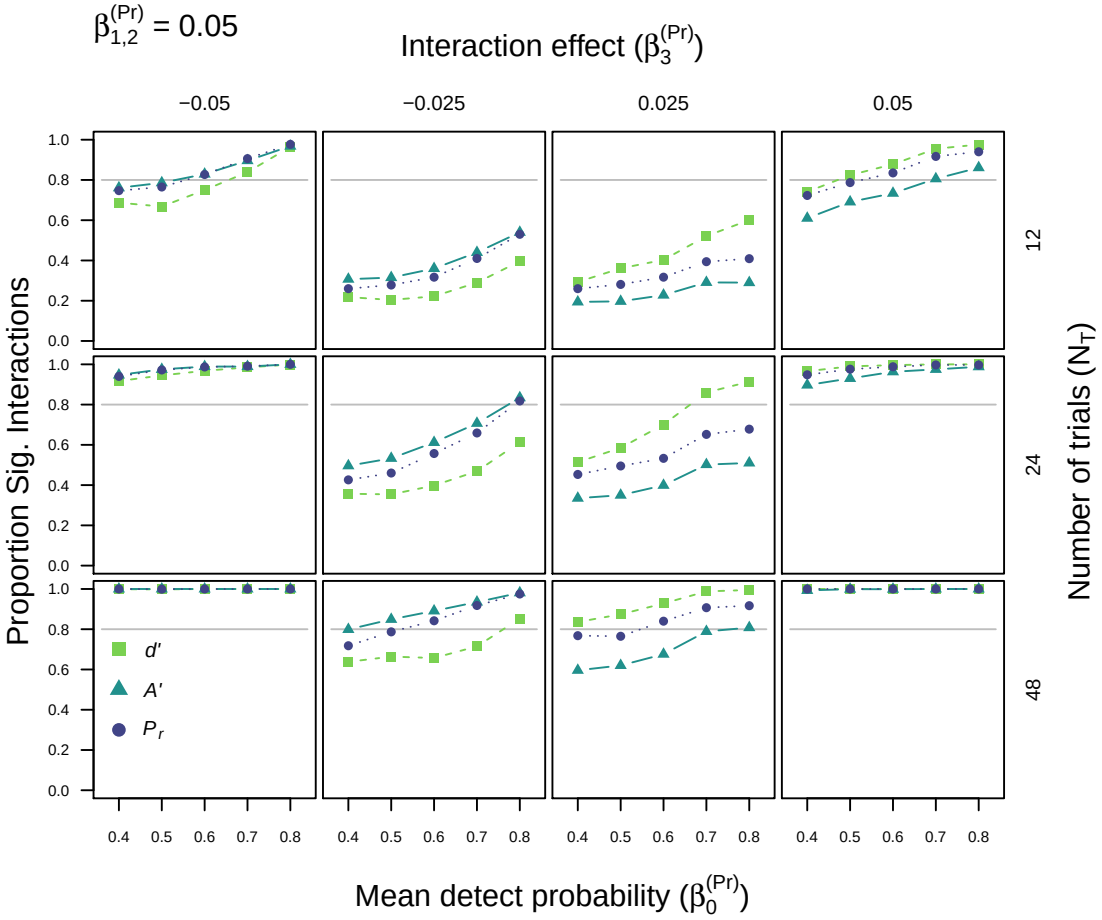

Figure 38: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.05.

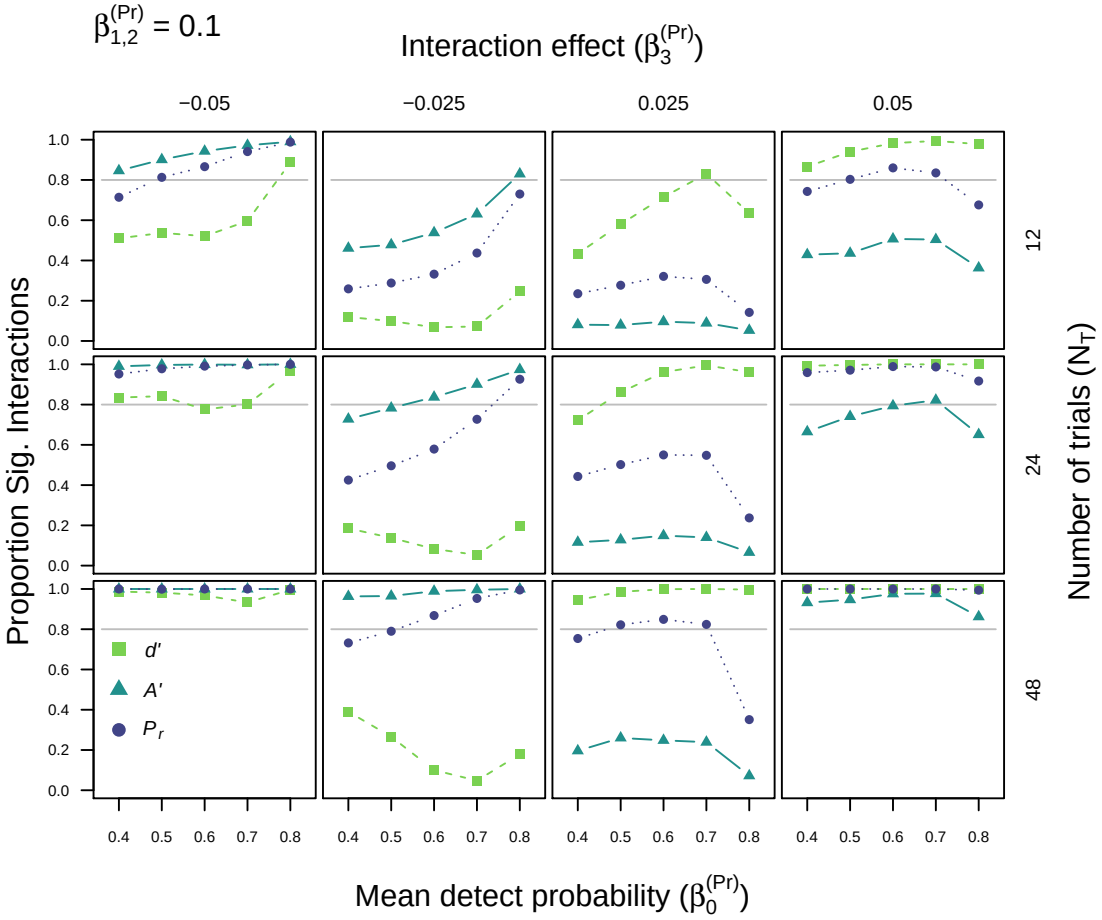

Figure 39: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.1.

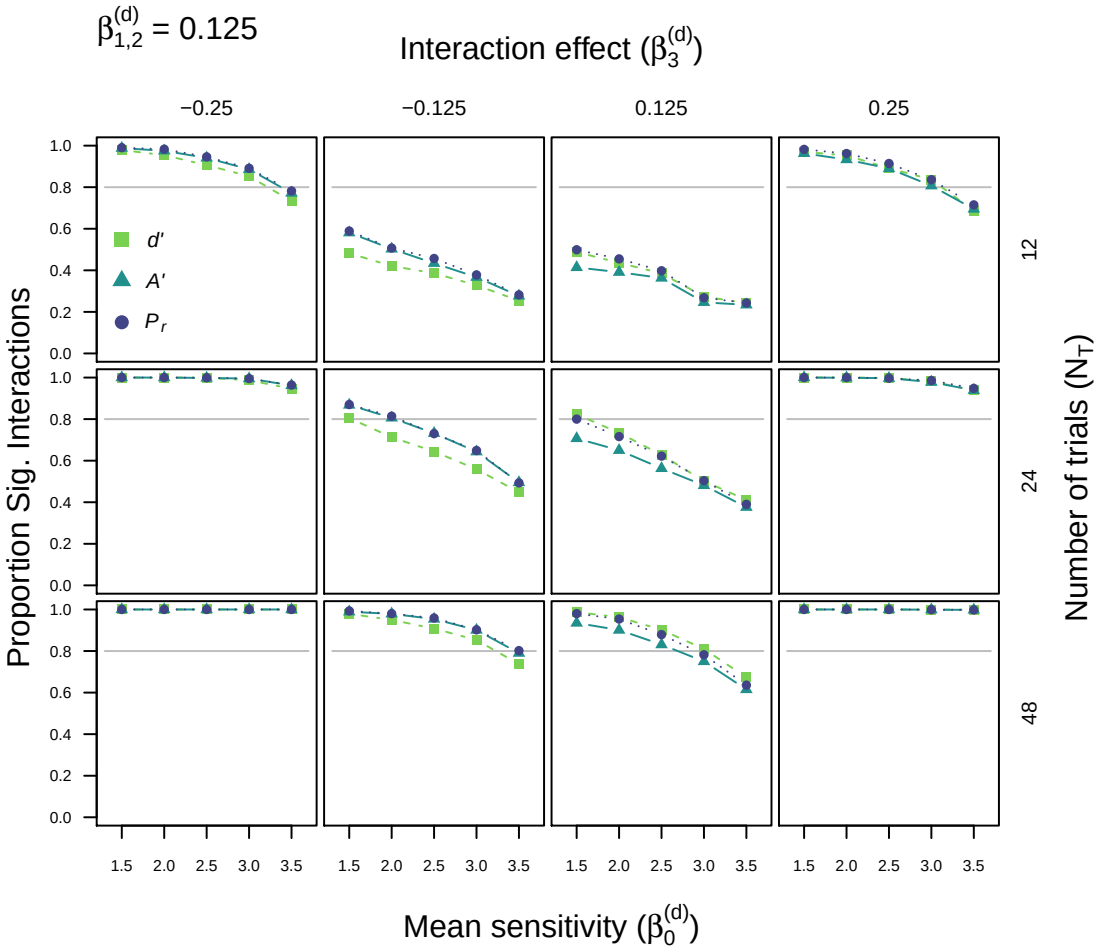

Figure 40: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

9.2 Signal detection

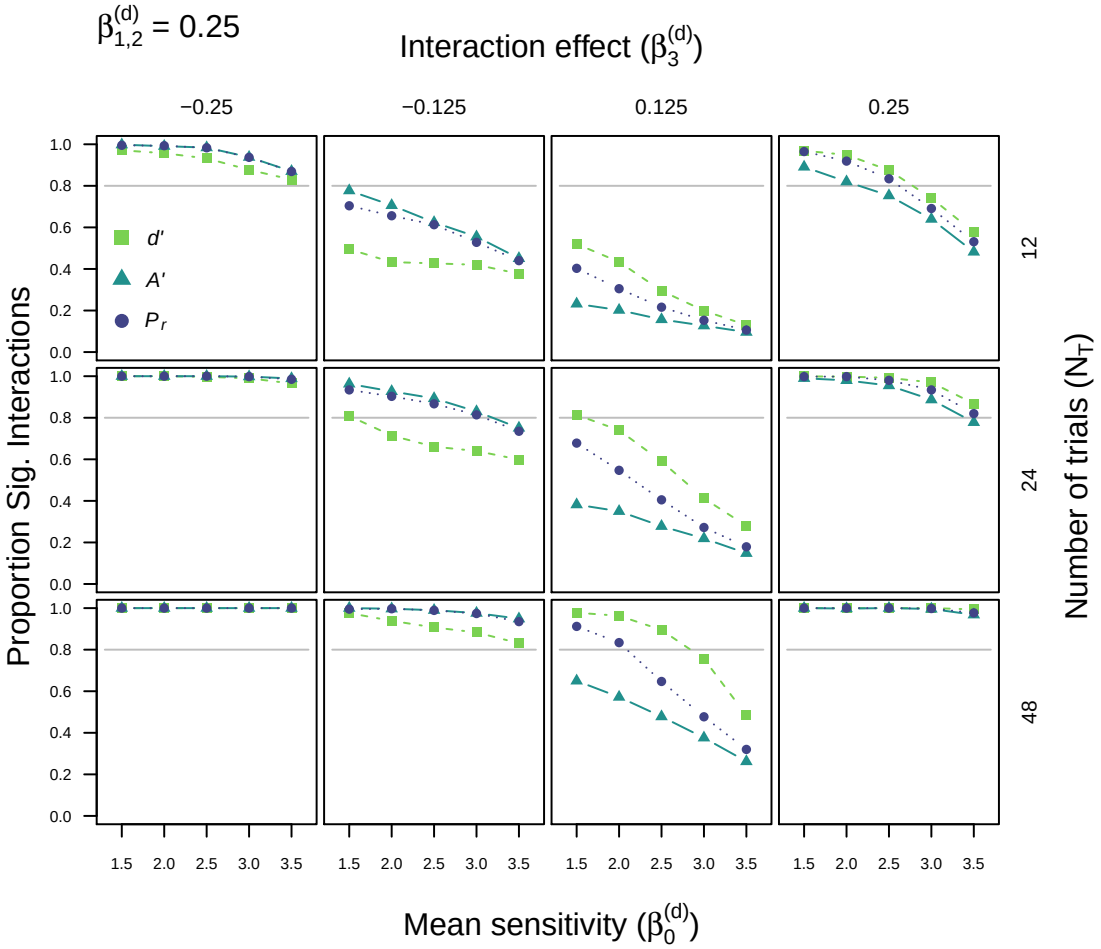

Figure 41: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

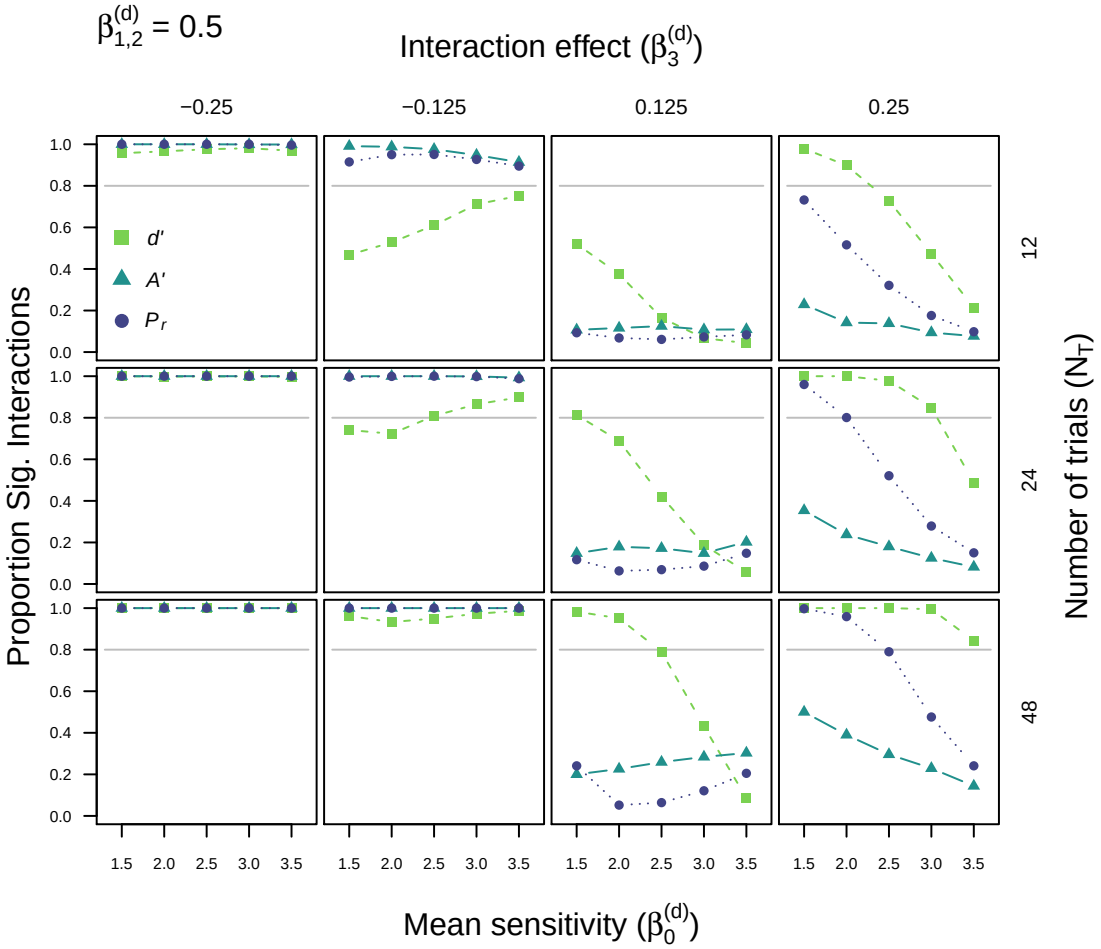

Figure 42: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

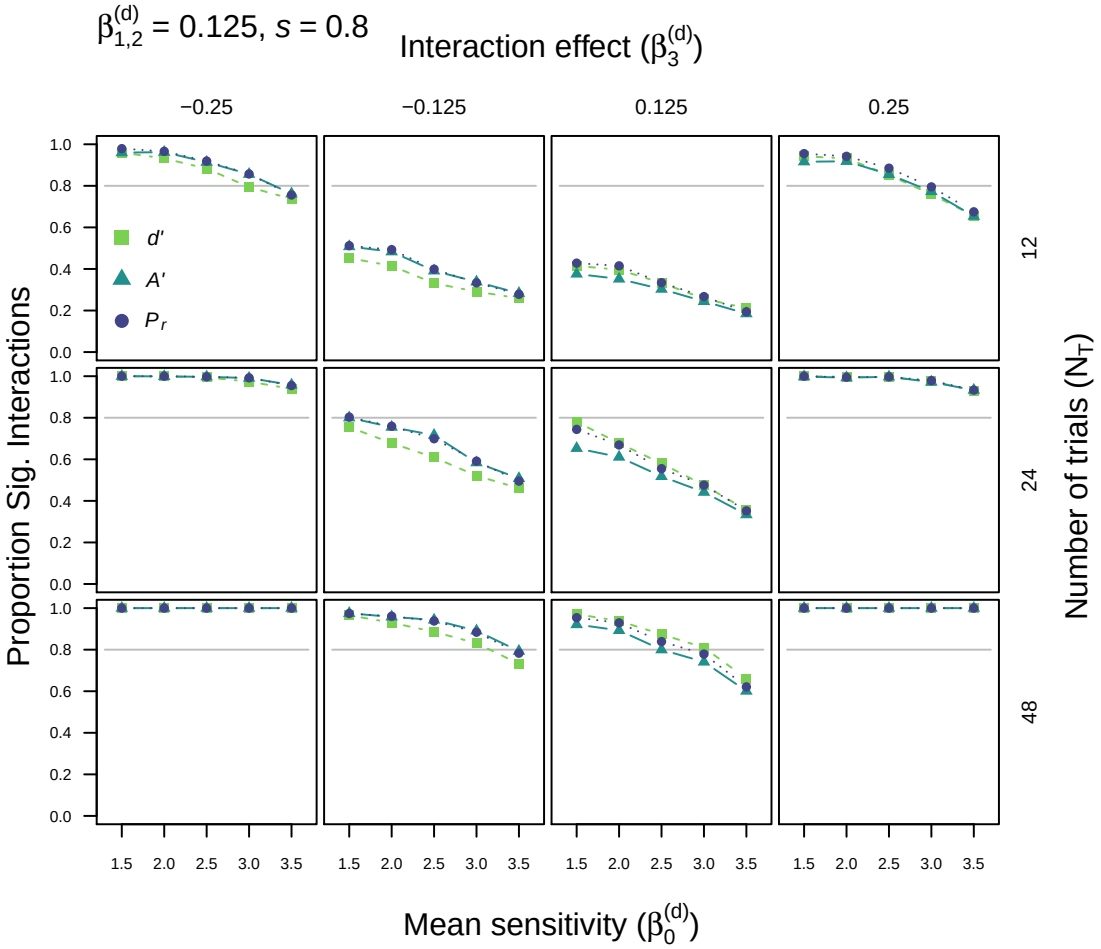

Figure 43: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

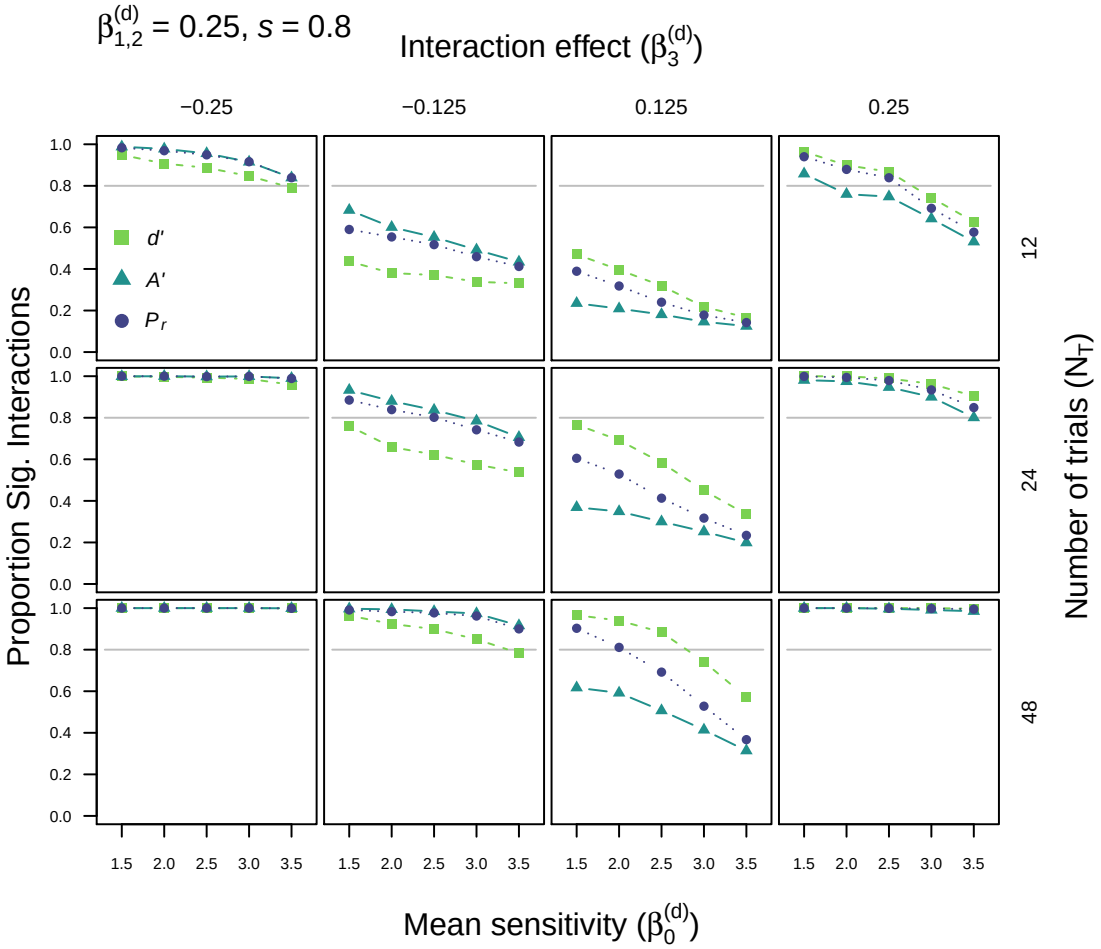

Figure 44: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

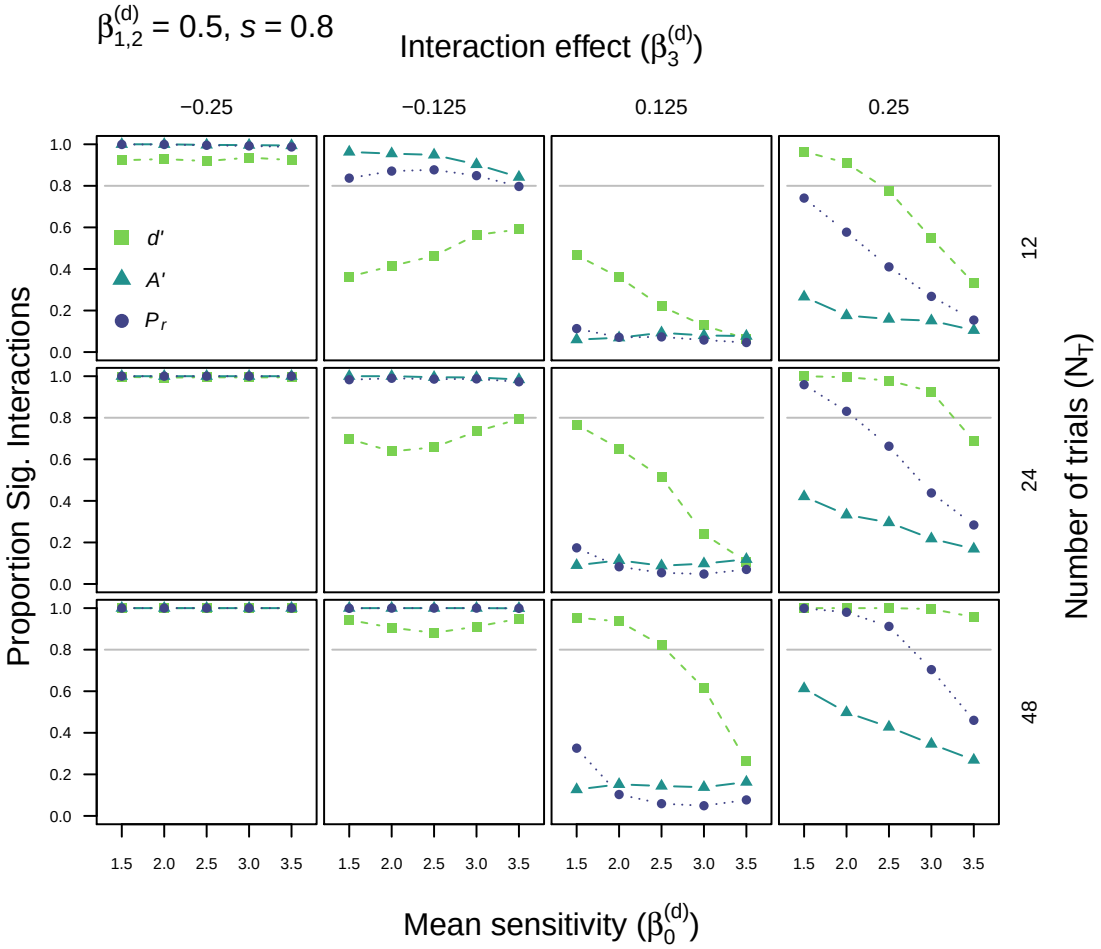

Figure 45: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

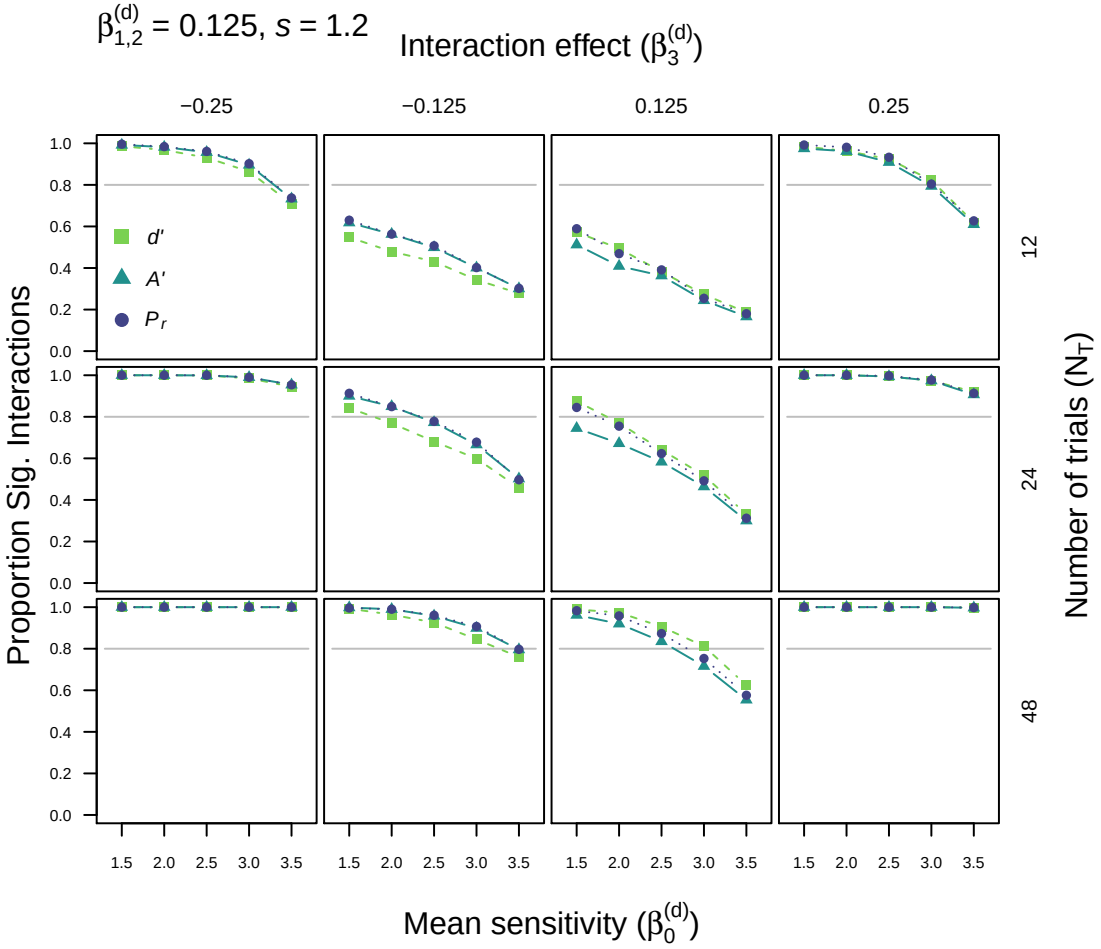

Figure 46: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

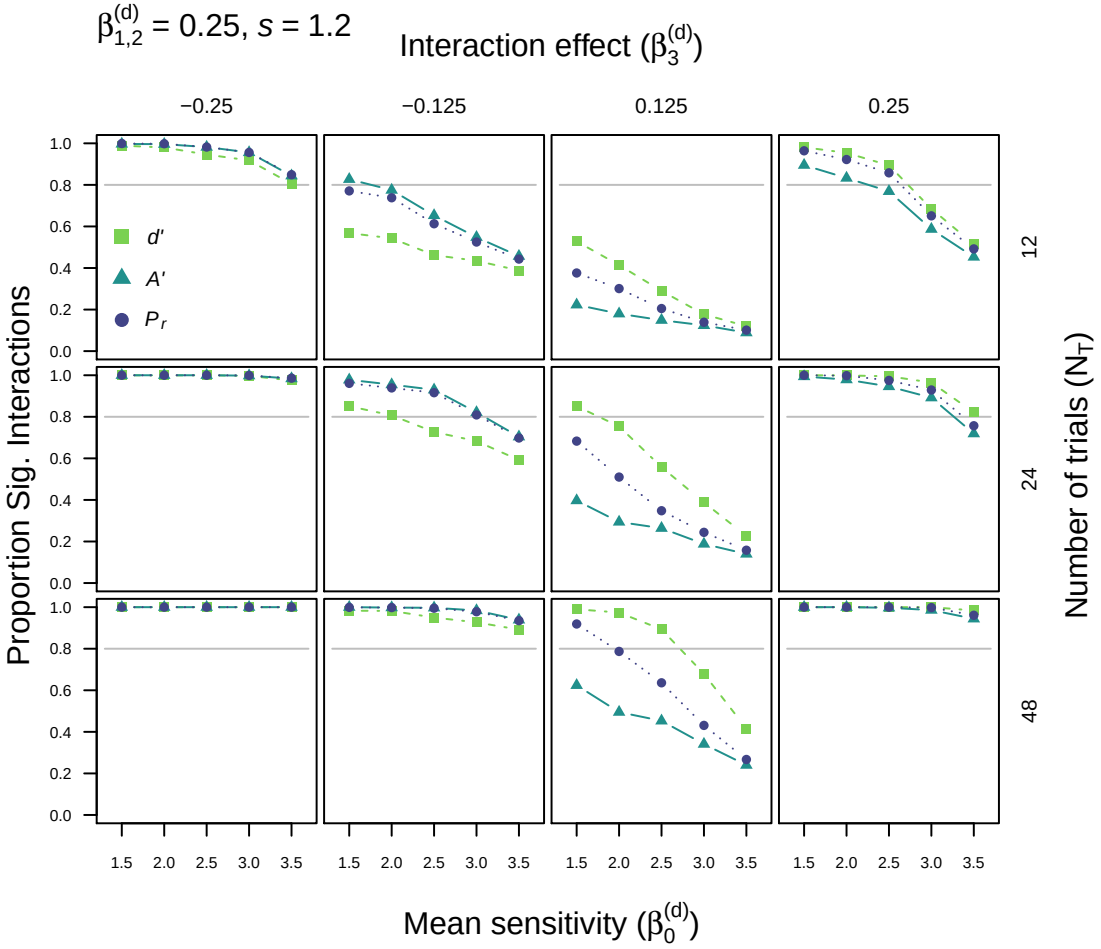

Figure 47: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

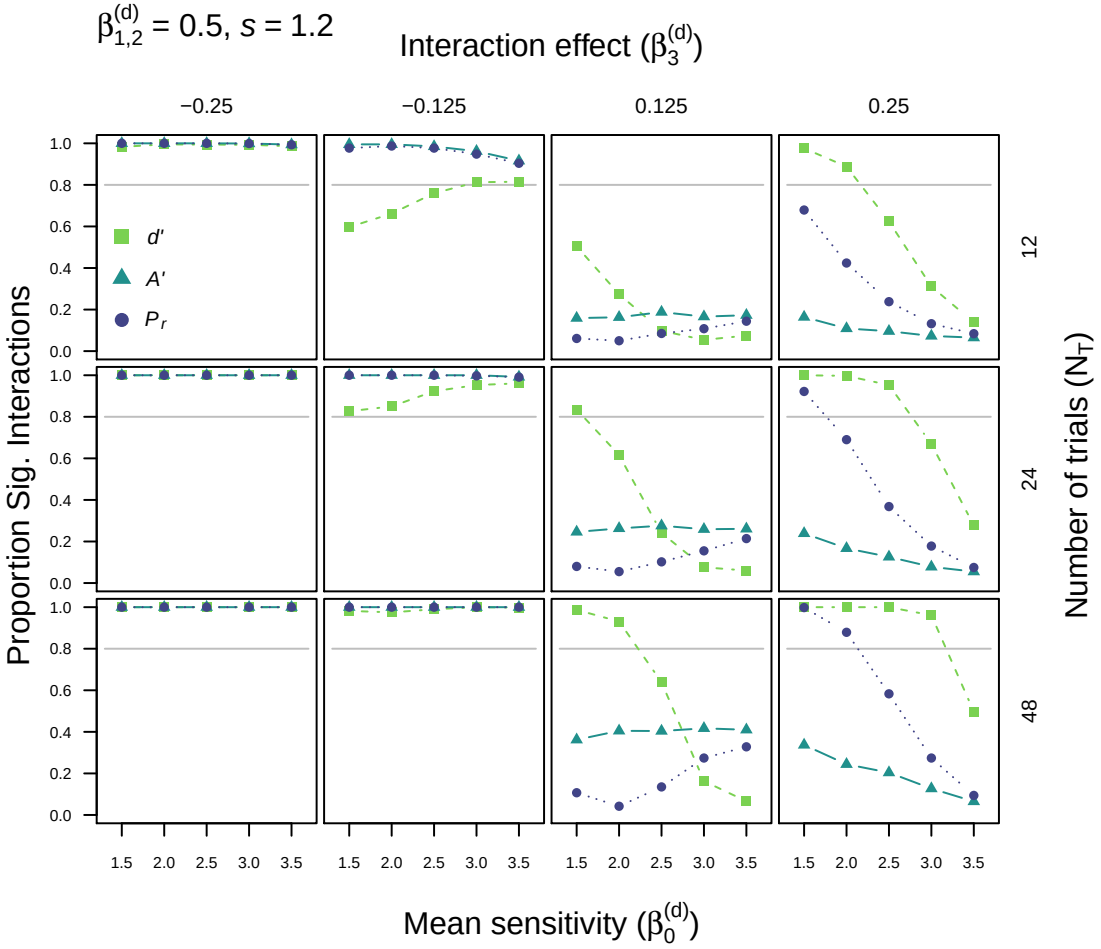

Figure 48: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and the number of trials per condition (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

10 Simulation 10: varying overall sensitivity and bias

In simulation 10 we fixed the number of trials per condition and subjects per group to 24 and added variation in bias. The grand mean bias was varied and individuals were allowed to randomly differ from this grand mean. Here we used the same parameter settings for overall and individual differences in bias as Simulation 4. The outcome of this tenth series of simulations can be found in Figures 49 to 60. As with the type I error simulations, there was little effect of introducing variability in bias. Thus we do not discuss these results in detail, as the next set of simulations produced more complex patterns

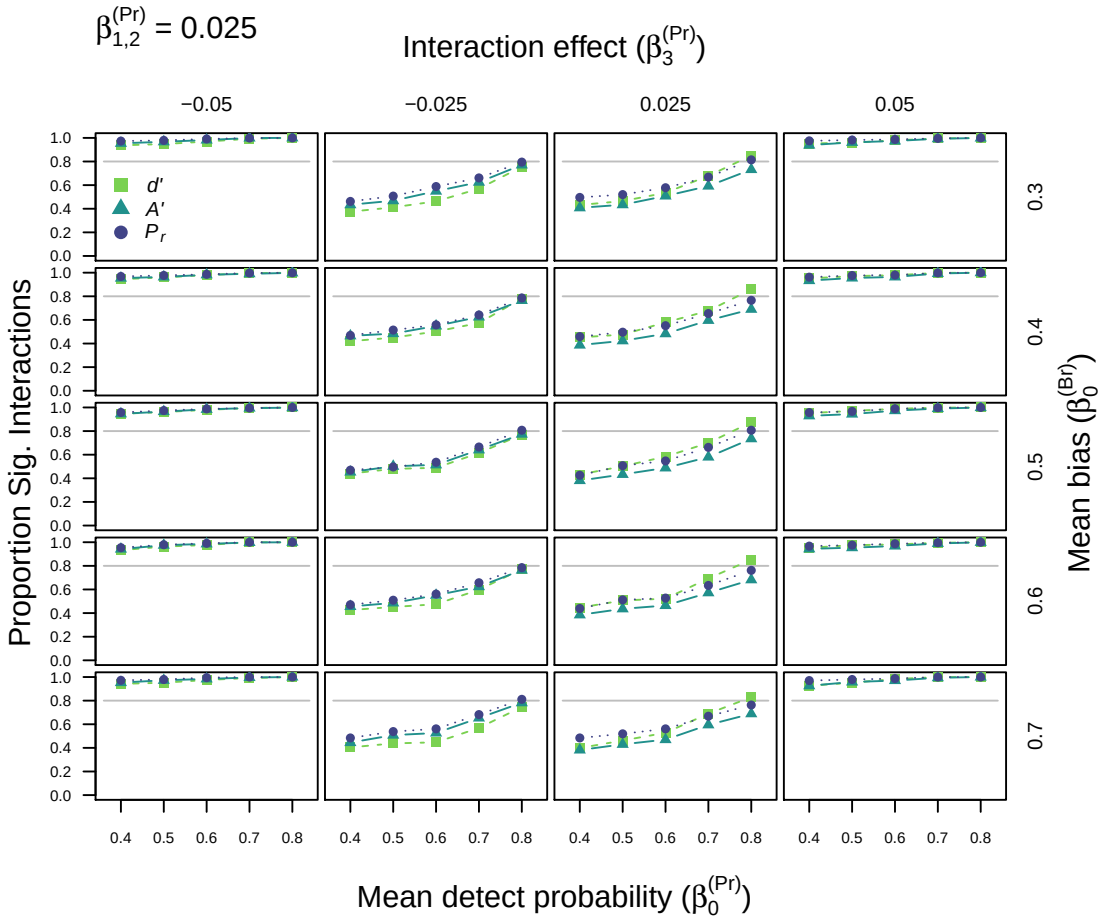

Figure 49: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the overall bias exhibited (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.025.

10.1 Two-high threshold

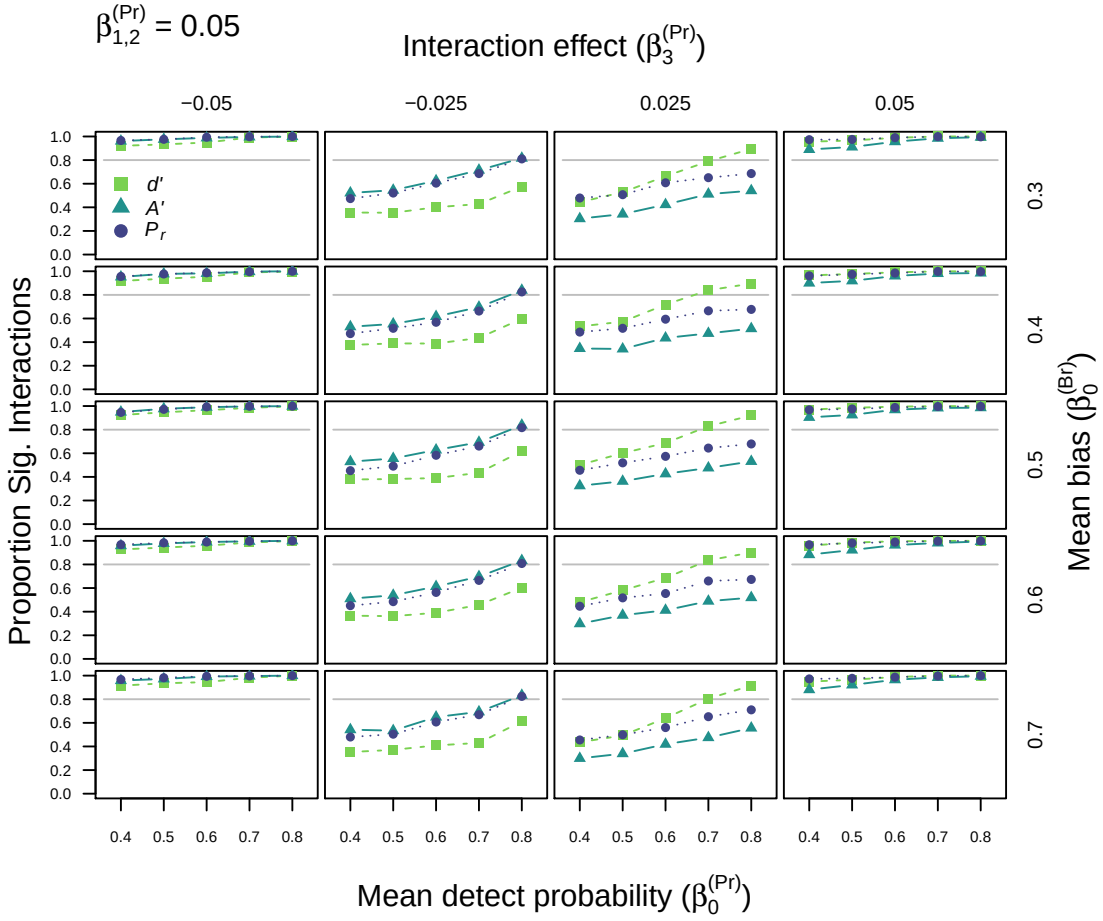

Figure 50: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the overall bias exhibited (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.05.

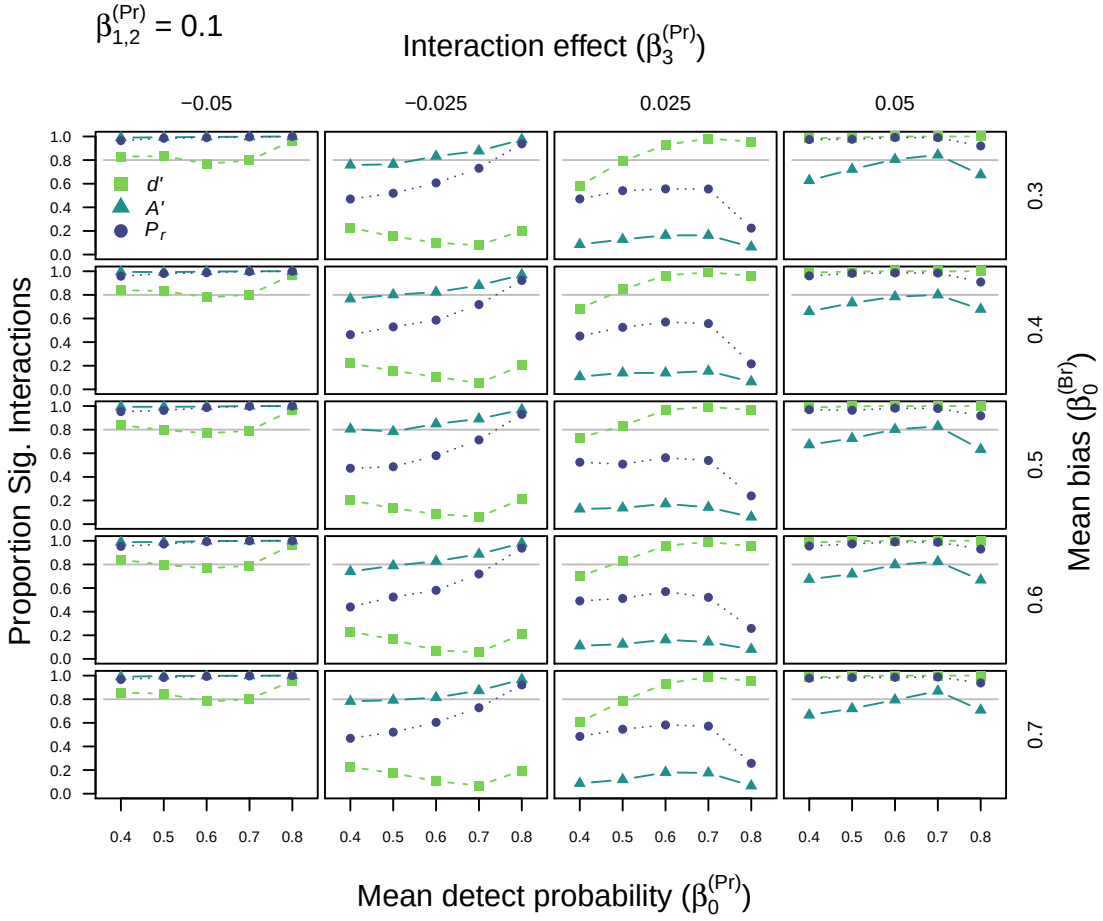

Figure 51: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall probability of detection (x axis), the magnitude and direction of interaction (left to right) and the overall bias exhibited (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.1.

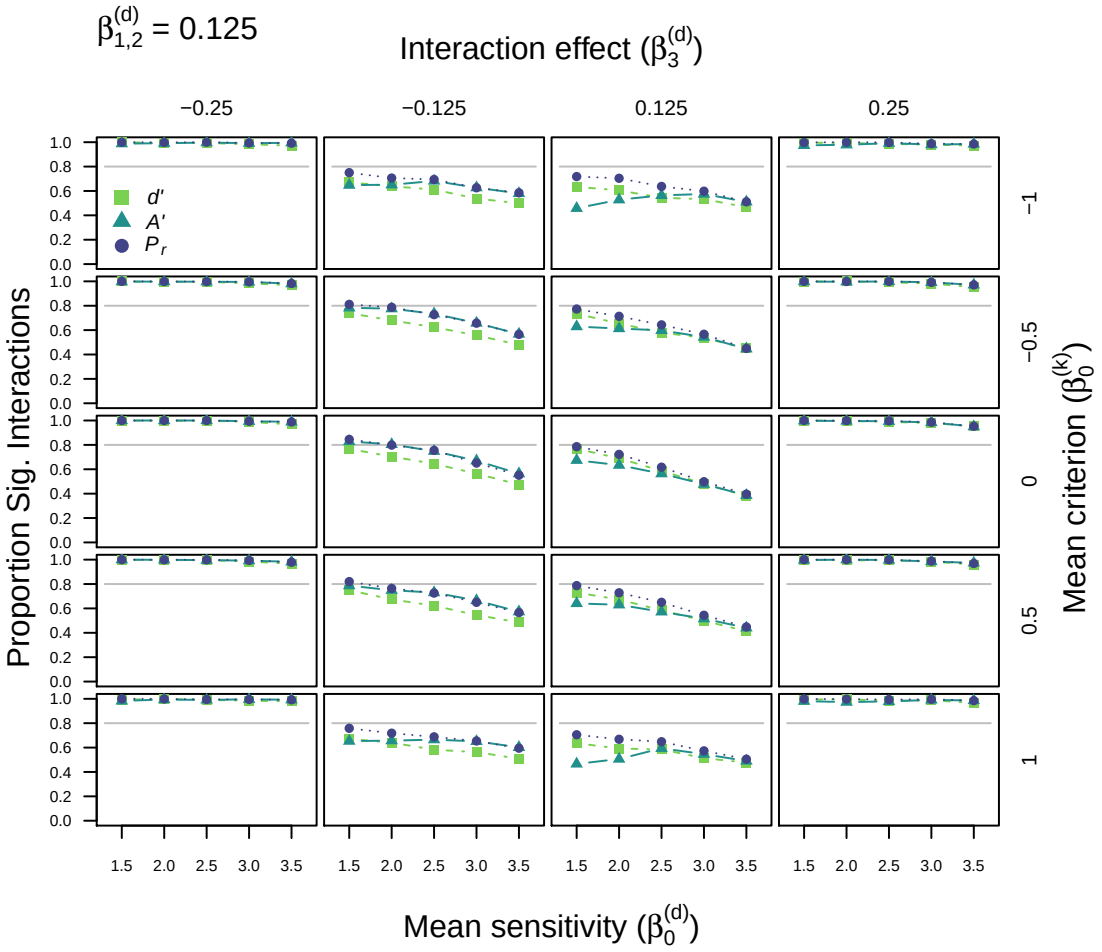

Figure 52: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

10.2 Signal detection

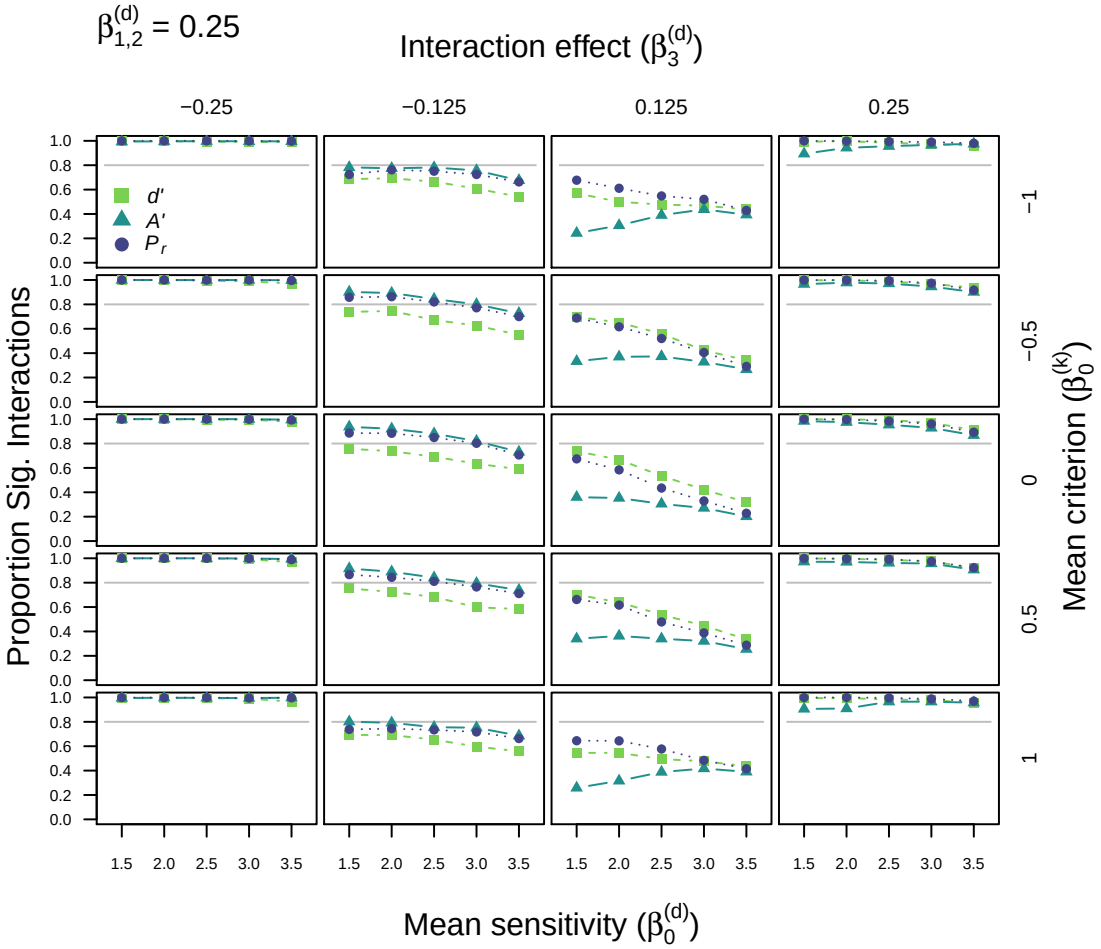

Figure 53: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

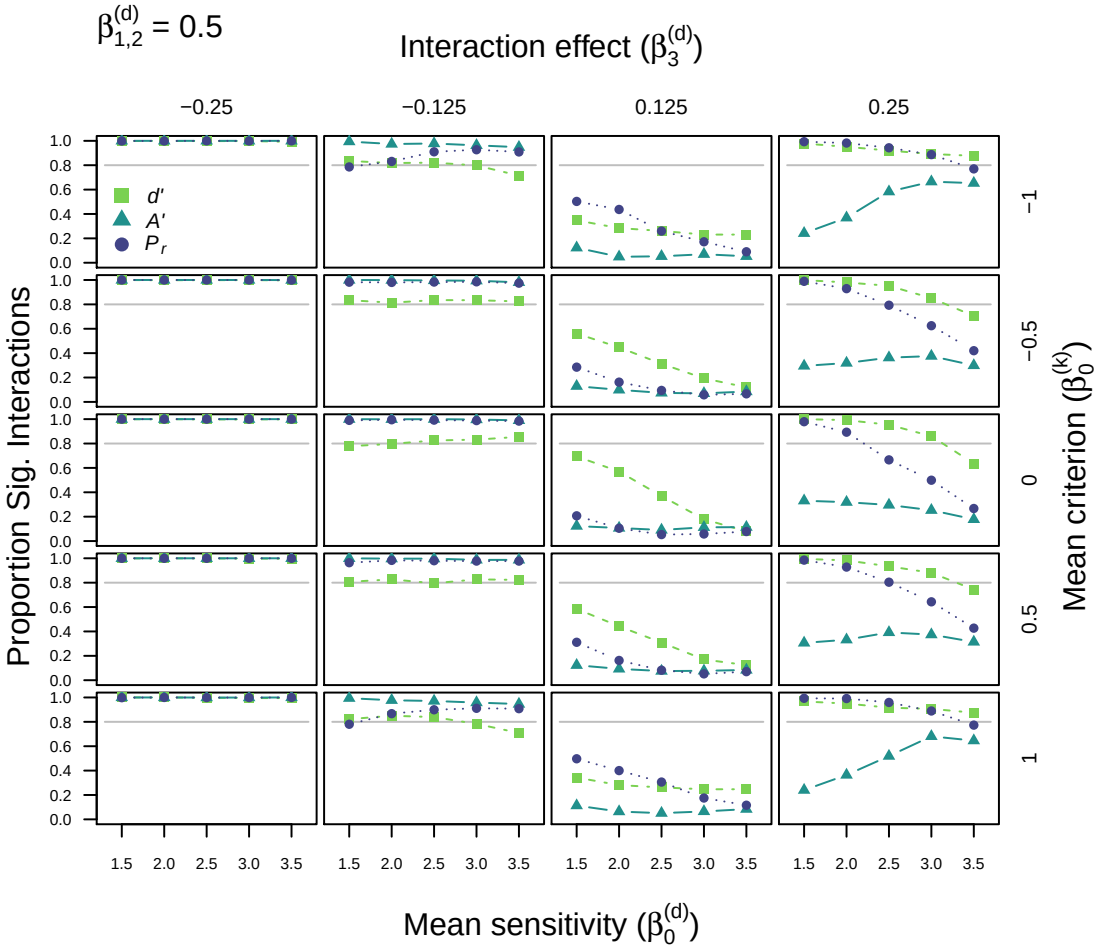

Figure 54: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

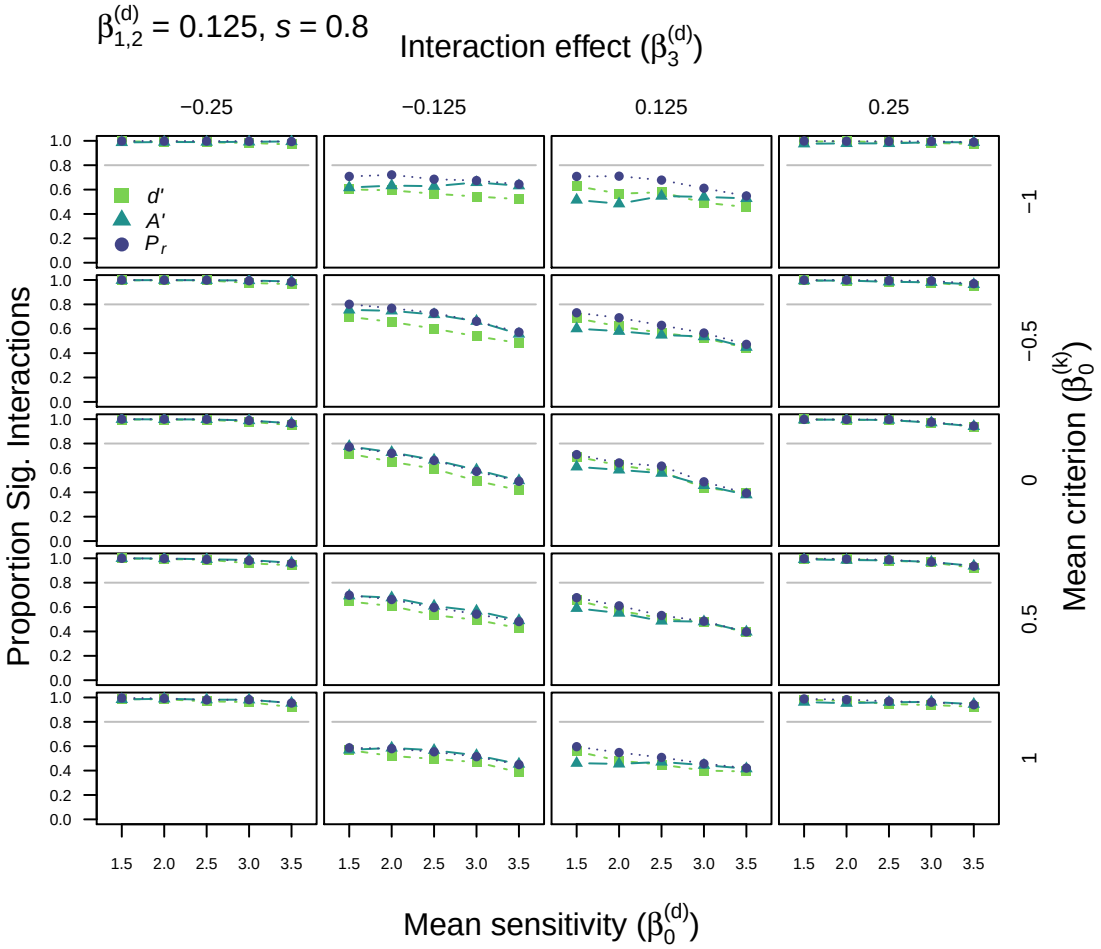

Figure 55: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

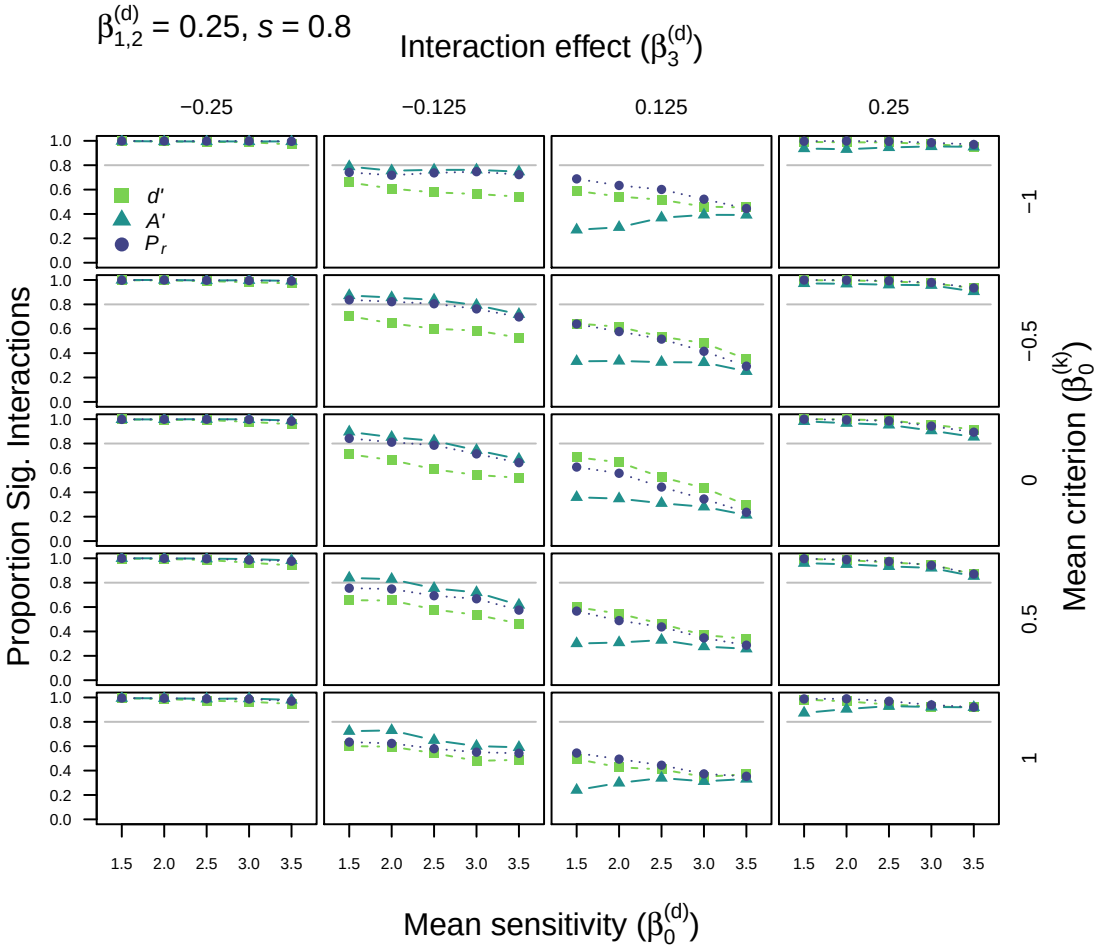

Figure 56: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

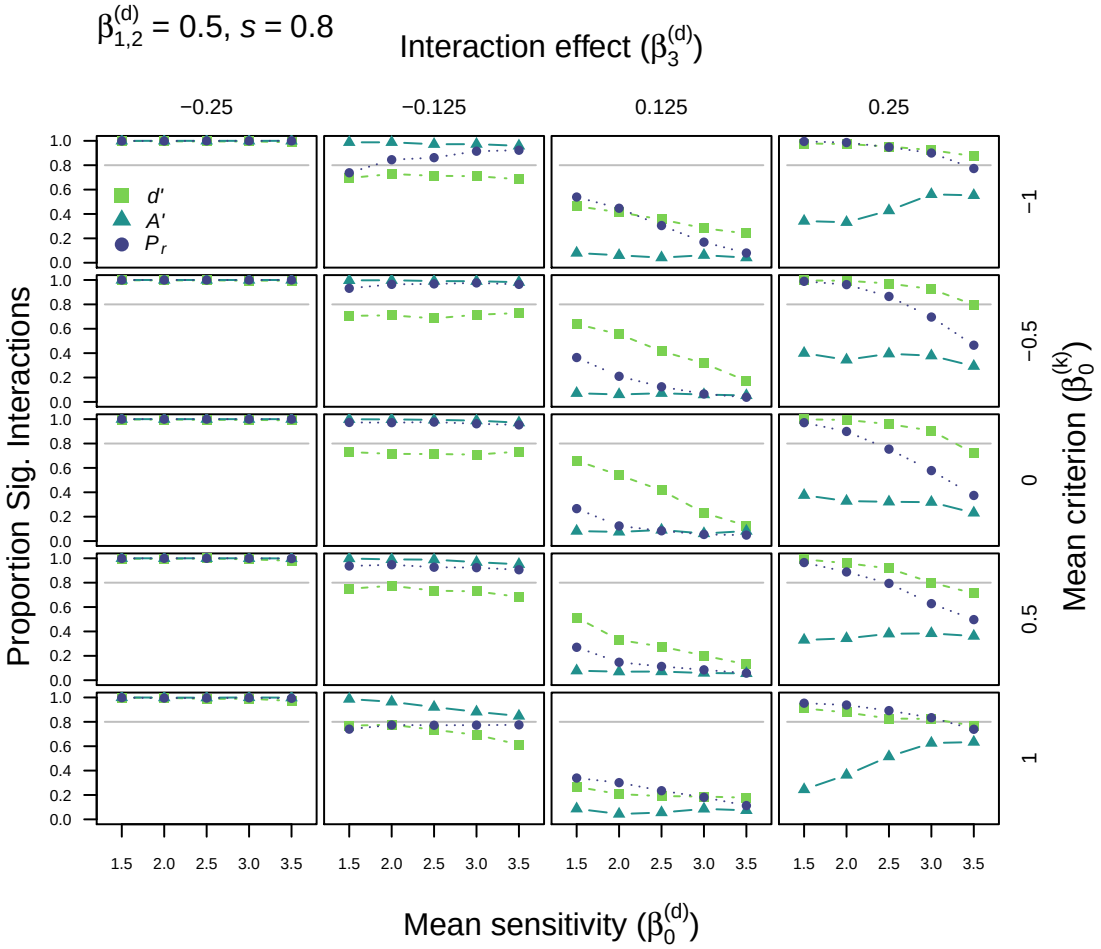

Figure 57: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

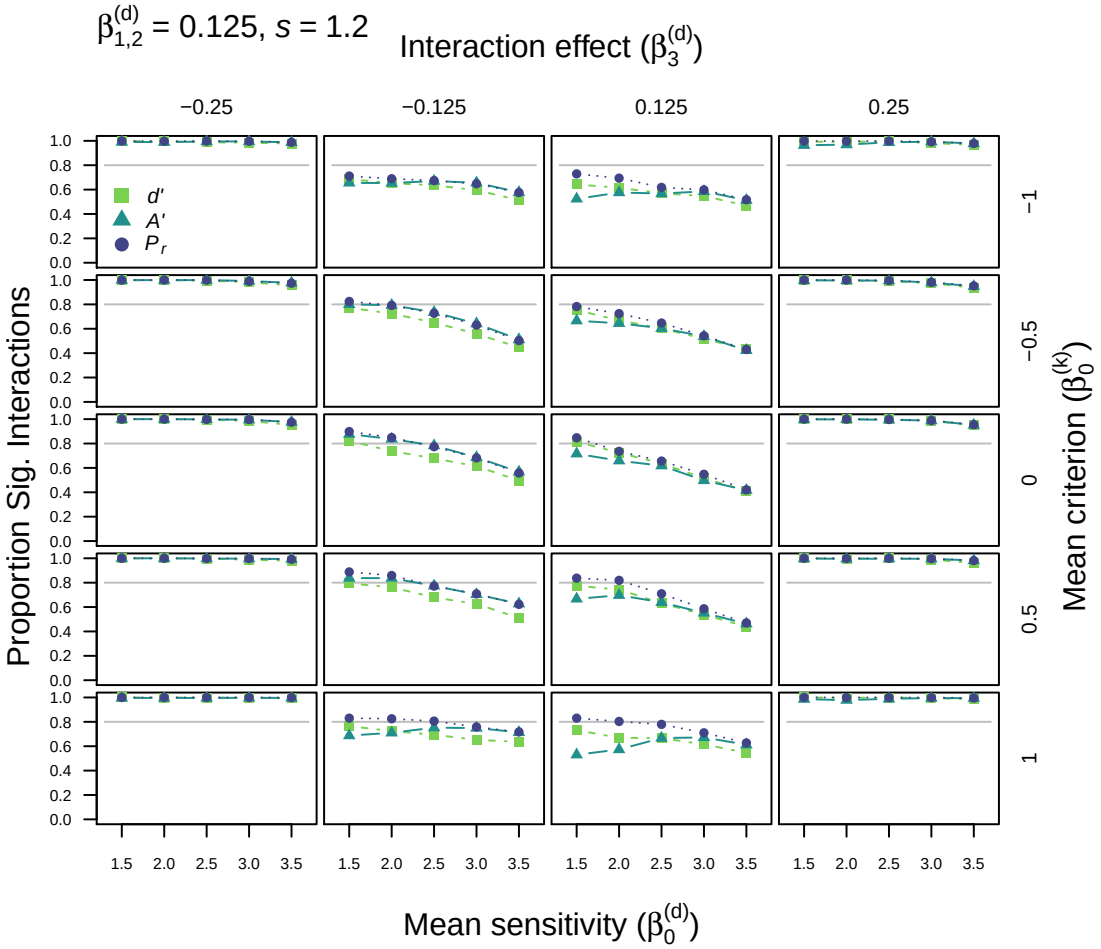

Figure 58: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

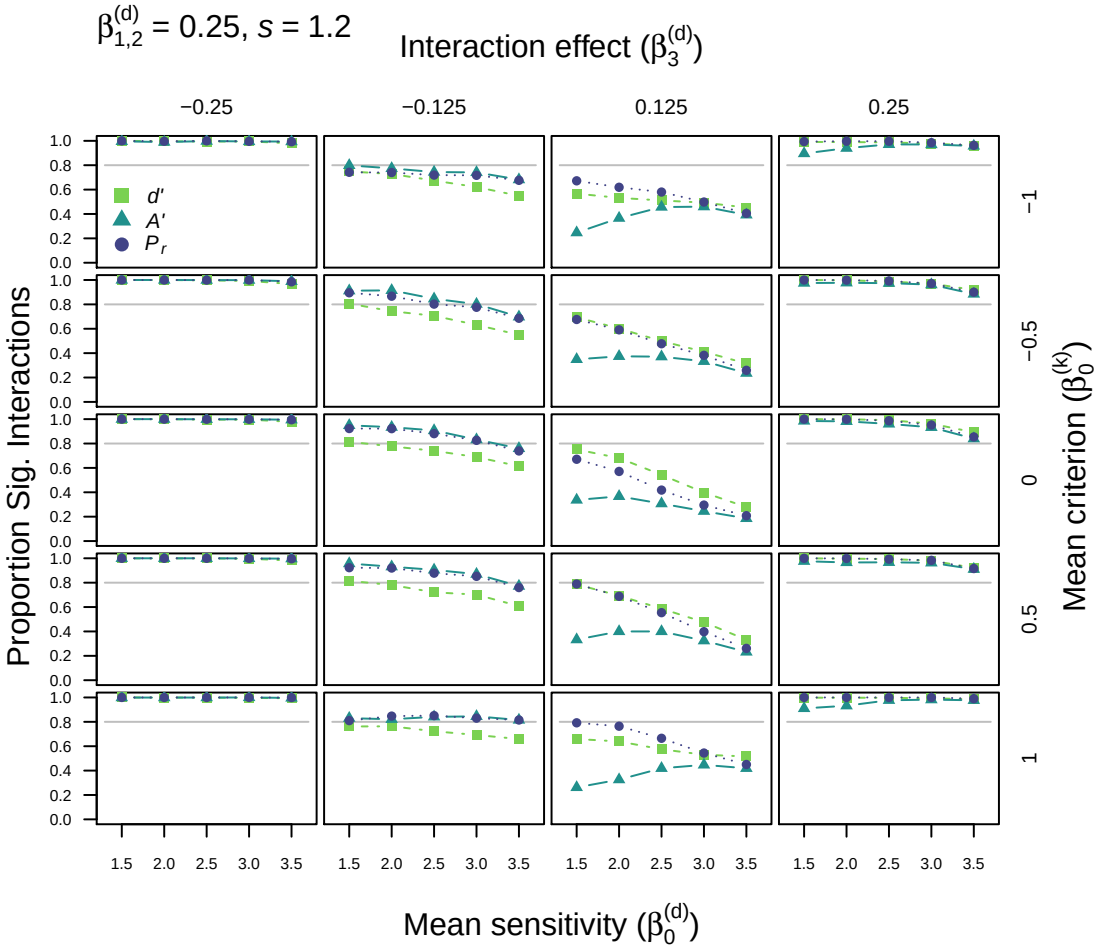

Figure 59: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

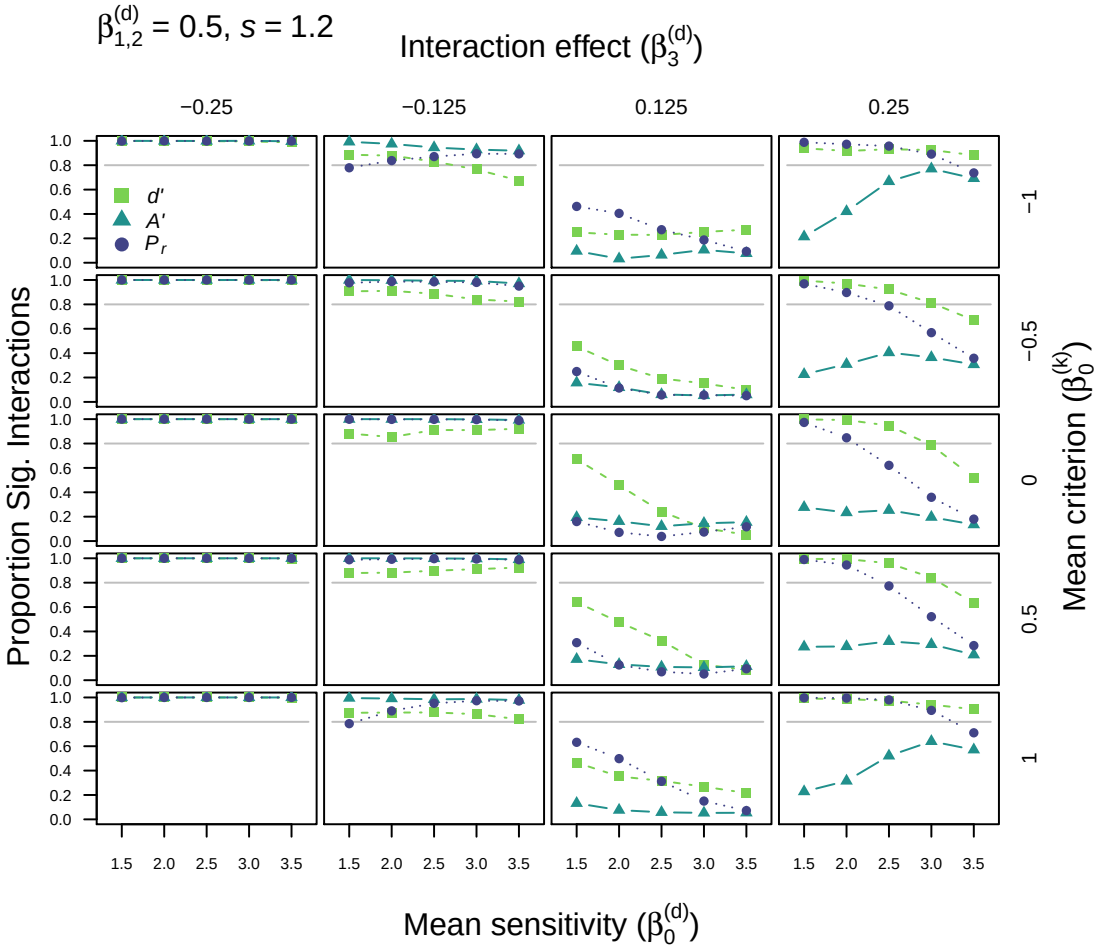

Figure 60: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall sensitivity (x axis), the magnitude and direction of interaction (left to right) and overall criterion placement (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

11 Simulation 11: introducing main effects on bias

In this simulation, we introduced main effects of both group and condition on response bias, in addition to the overall variation and individual differences in bias added in the previous simulation. Relative to simulation 10, these simulations produced a complex pattern of results so we discuss them in detail below.

11.1 Two-high threshold

In this set of simulations overall bias ($\beta_0^{B_r}$) was varied from 0.3 to 0.7 in steps of 0.1 and main effects of group and condition on bias ($\beta_1^{B_r} = \beta_2^{B_r}$) were set to either 0, 0.025, 0.05, or 0.1. As we were examining power for tests of discriminability the interaction effect ($\beta_3^{P_r}$) was varied through -0.05, -0.025, 0.025, and 0.05. Once again we also examined power under three magnitudes of main effect on detection ($\beta_1^{P_r} = \beta_2^{P_r}$) and overall discriminability ($\beta_0^{P_r}$) was fixed to 0.6. Figures 61, 62, and 63 present the results of simulations with small, medium, and large main effects, respectively.

Figure 61 shows clear differences between the measures arose even with small main effects on detection (0.025). When main effects on response bias were large (0.05 or 0.1), d' was clearly under-powered relative to P_r for under-additive interactions (negative interaction coefficients), whereas it was clearly over-powered for over-additive interactions (positive interaction coefficients). A similar pattern was present for A', but somewhat less pronounced. These trends are exacerbated by increasing the size of main effects on detection (Figure 62).

Figure 63 shows that even for larger under-additive interactions, d' is more likely than other measures to produce a miss. This tendency increases with the size of main effects on bias and becomes increasingly dependent on the overall (grand mean) level of bias exhibited (x-axis). d' is particularly under-powered for under-additive interactions when the overall bias is conservative (left panels). When considering over-additive interactions, A' is clearly at a disadvantage relative to the other measures and especially so for more liberal overall guessing biases (right panels). The reason behind this pattern, as before, lies with the shape of the predicted ROC functions for each of these measures (see main manuscript for discussion). With the two-high threshold generative model, variation in bias merely shifts the (f, h) points relative to the negative diagonal but does not change their placement relative to the chance line (positive diagonal). d' and A', on the other hand, predict a curved relationship between f and h as bias varies. Had we also varied the overall sensitivity of observers, the disagreement between measures would have been even more pronounced at high levels as shown in the previous simulations.

For a large under-additive interaction (far left panels of Figure 63) the tendency of d' to distort high threshold data into over-additive interactions is not strong enough and interactions still appear with this measure. This is true until there are large main effects on bias and a liberal overall bias, in which case false-alarm rates are quite low, reducing d's ability to detect the interaction. A similar pattern is present for A' at small under-additive interactions and this pattern flips when the

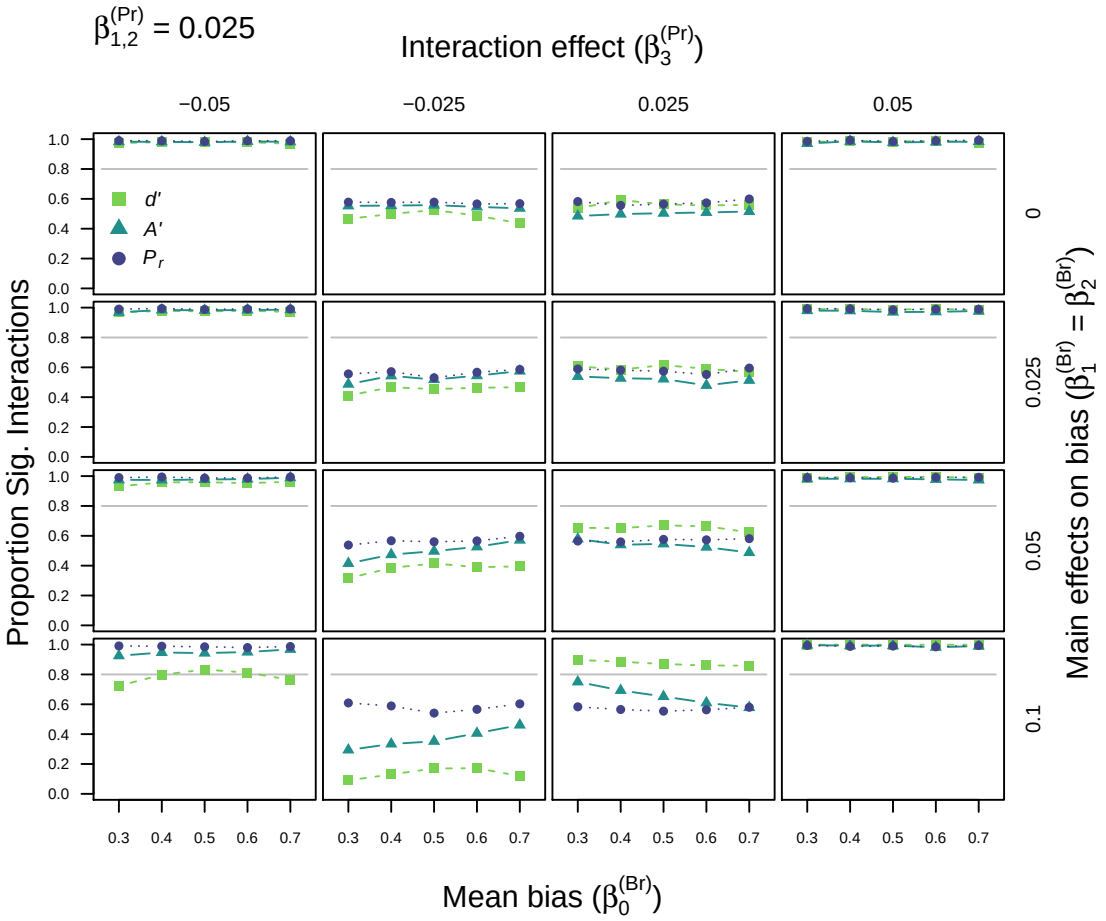

Figure 61: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall guessing bias (x axis), the magnitude and direction of interaction (left to right) and the magnitude of main effects on bias (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.025.

interactions are over-additive. In this case, d' has power of essentially 1 regardless of bias parameters, given that the true interaction is combined with the distorting effect of applying this measure to two-high threshold data. As noted in the manuscript, being *overpowered* in this case is not a good thing as it results from a distortion of the true effect size.

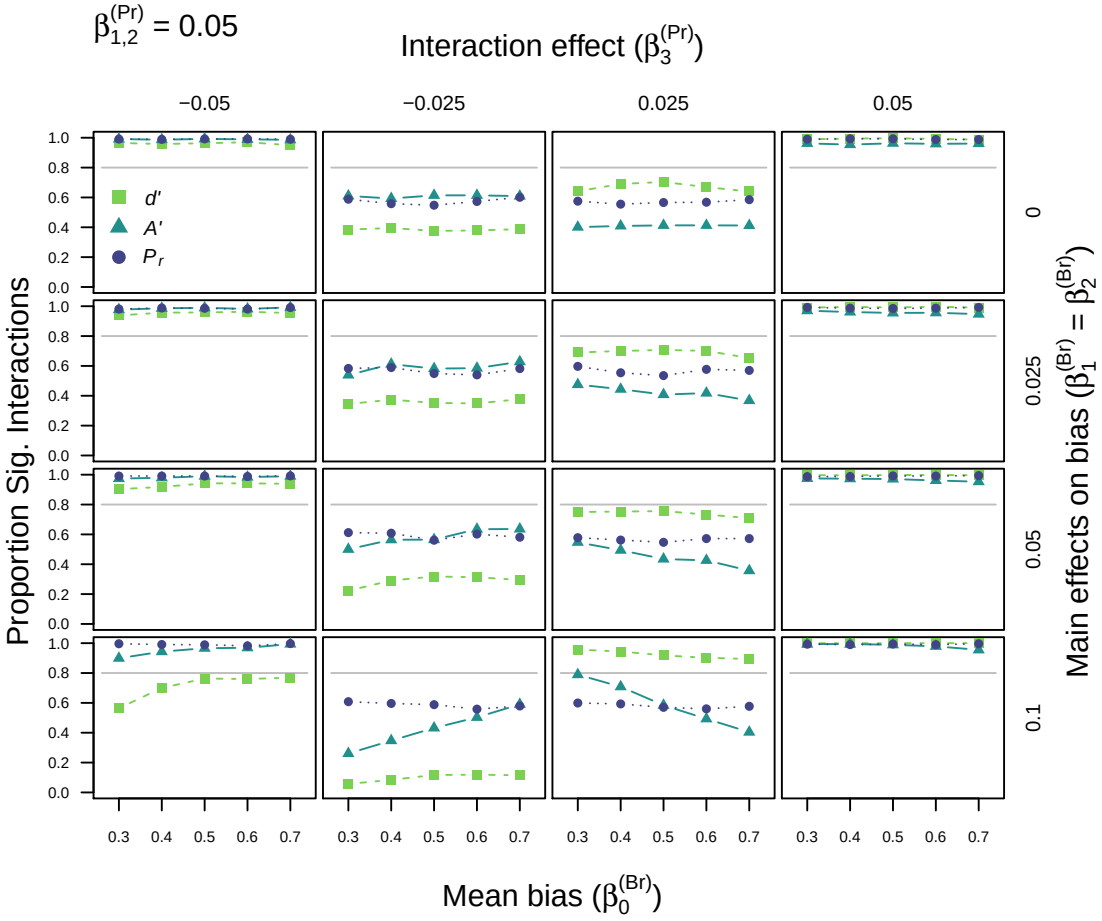

Figure 62: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall guessing bias (x axis), the magnitude and direction of interaction (left to right) and the magnitude of main effects on bias (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.05.

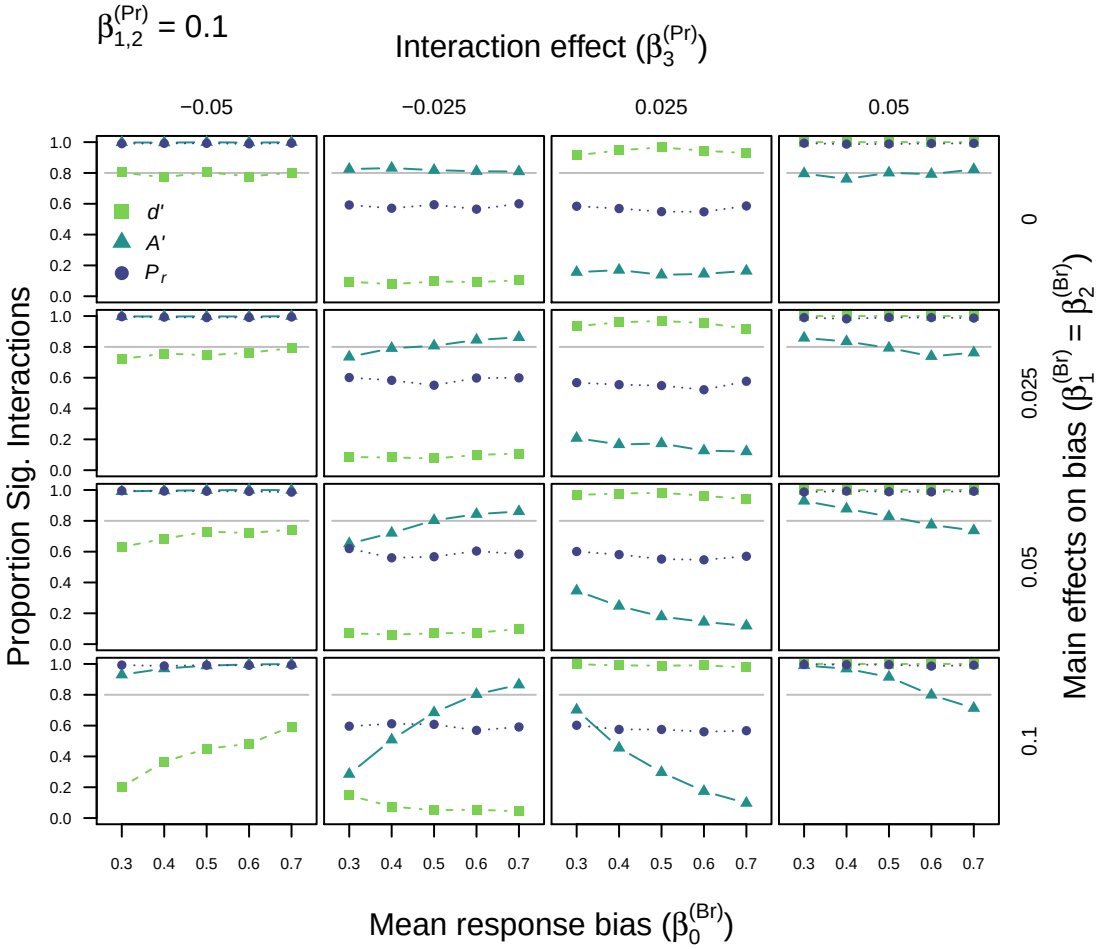

Figure 63: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall bias (x axis), the magnitude and direction of interaction (left to right) and the magnitude of main effects of group and condition on bias (top to bottom). Note: this figure depicts the case where main effects on detection equal 0.1.

11.2 Signal detection

For the signal detection simulations, grand mean sensitivity (β_0^d) was fixed to 2 and main effects on sensitivity ($\beta_1^d = \beta_2^d$) were varied through 0.125, 0.25, and 0.5. Overall grand mean criterion placement (β_0^k) was varied from liberal, -1, to conservative, 1, in steps of 0.5 and the main effects on criterion ($\beta_1^k = \beta_2^k$) could take on the values, 0, 0.125, 0.25, or 0.5. The crucial interaction on sensitivity (β_3^d) could be -0.25, -0.125, 0.125, or 0.25 for under- and over-additive effects, respectively.

The pattern of results is rather complicated for these simulations, as shown in Figure 66. For under-additive interactions all measures perform fairly well. Power for d' and P_r drops at extreme liberal overall criterion placements due to high hit rates across the board, making false-alarm rate the deciding factor. For over-additive interactions (the right-hand panels of Figure 66) estimates of power appear rather erratic; we attempt to briefly summarize the reasons for this.

As shown in Simulation 7, when there is a small over-additive interaction on d and overall bias is neutral, P_r is greatly under-powered relative to d'. However, as shown in the third position of the top row of Figure 66, varying overall bias improves P_r's power while reducing power for d'. This is because moving along the Gaussian signal detection ROCs, with no main effects on criterion, accentuates the disparity between the highest performing group or condition in ROC space producing an over-additive interaction for P_r. Simultaneously, the ROC points congregate close to $h = 1$ or $f = 0$ for liberal and conservative biases, respectively, reducing power for d' as only one of the rates comes to dominate inference. Adding in main effects on criterion accentuates these trends and also produces an asymmetry as the most sensitive group/ condition is also the most liberal (see manuscript for discussion), therefore their false-alarm rate is essentially zero and inference becomes increasingly dependent on hit rate as overall bias becomes more liberal, reducing d's power further.

Perhaps the strangest pattern is seen for the largest over-additive interaction when there are two large main effects on criterion placement (see bottom right panel of Figure 66). In this case power for P_r starts quite high with an overall liberal criterion placement (-1) then drops dramatically as grand mean criterion becomes more conservative (-0.5). Power then raises again, to just under 60%, when overall bias is neutral, then drops for conservative responding. This, once again, is due to the curvature of the signal detection ROCs. At very liberal overall bias the points are extremely spread out in ROC space with group 2 in condition 2 achieving very high hits and low false-alarms, producing a clear over-additive interaction for P_r (and d'). At a slightly less liberal bias, the distance between the (f, h) pairs is reduced and

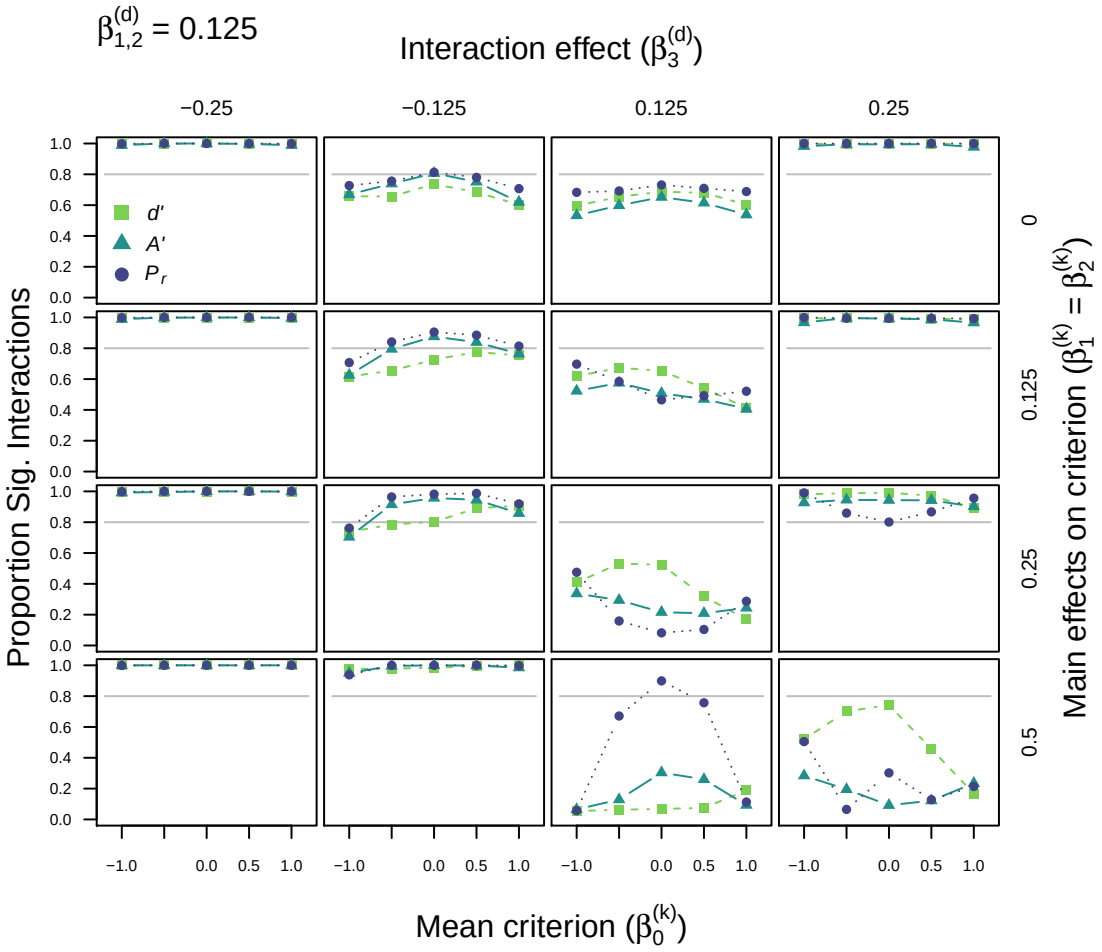

Figure 64: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

they are relatively evenly spaced from the positive diagonal, resulting in only main effects for P_r. As overall criterion placement moves through neutral to conservative, an under-additive trend appears in P_r, increasing power for interactions once again. These complicated patterns are generally not present for A', which has particularly poor power for over-additive interactions in GEV-SDT derived data.

In summary, when data conform to the expectation of signal detection theory, P_r will tend to miss over-additive interactions, whereas for data conforming to two-high threshold theory, d' has a tendency to miss under-additive interactions. This is complicated further when groups and conditions differ in their response biases, especially for signal detection data, in which extreme bias leads to ceiling/ floor effects in h or f, resulting in poor performance even for the principled measure, d' (see manuscript for discussion).

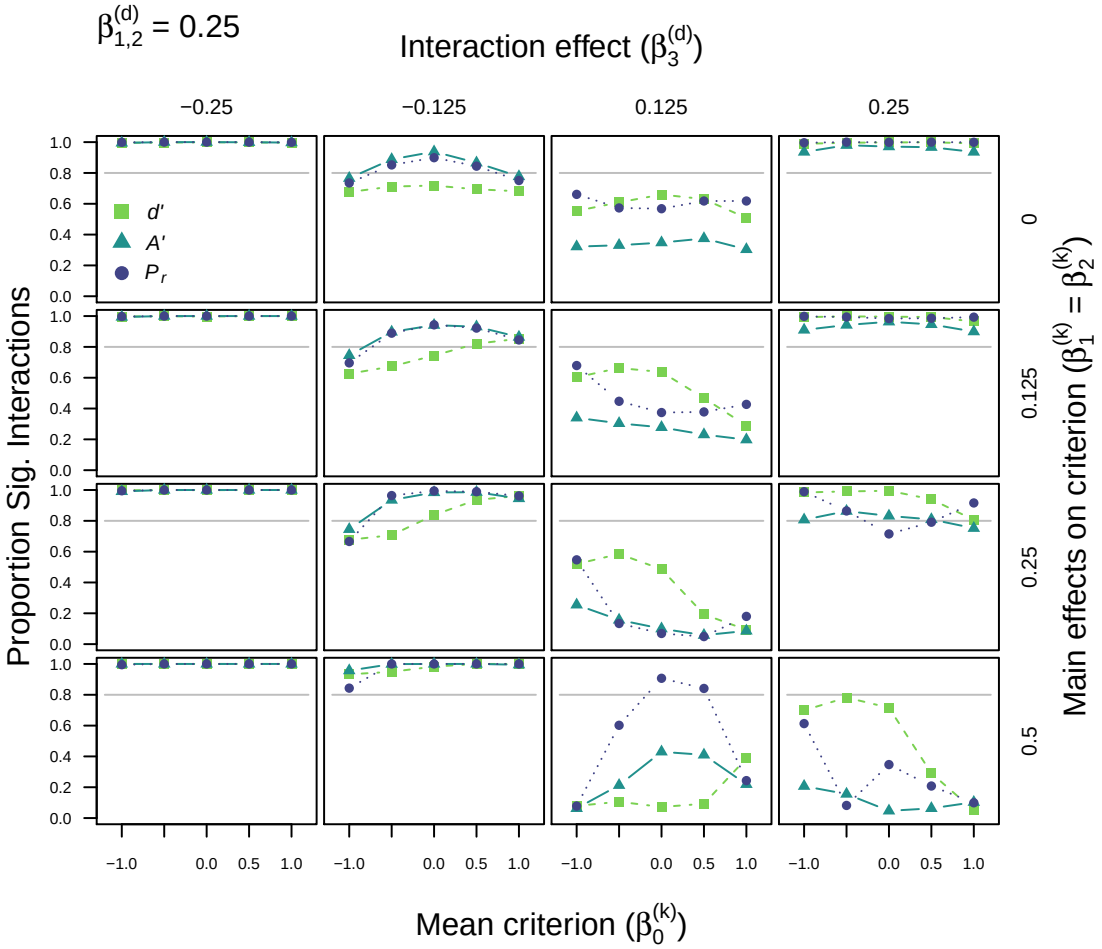

Figure 65: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

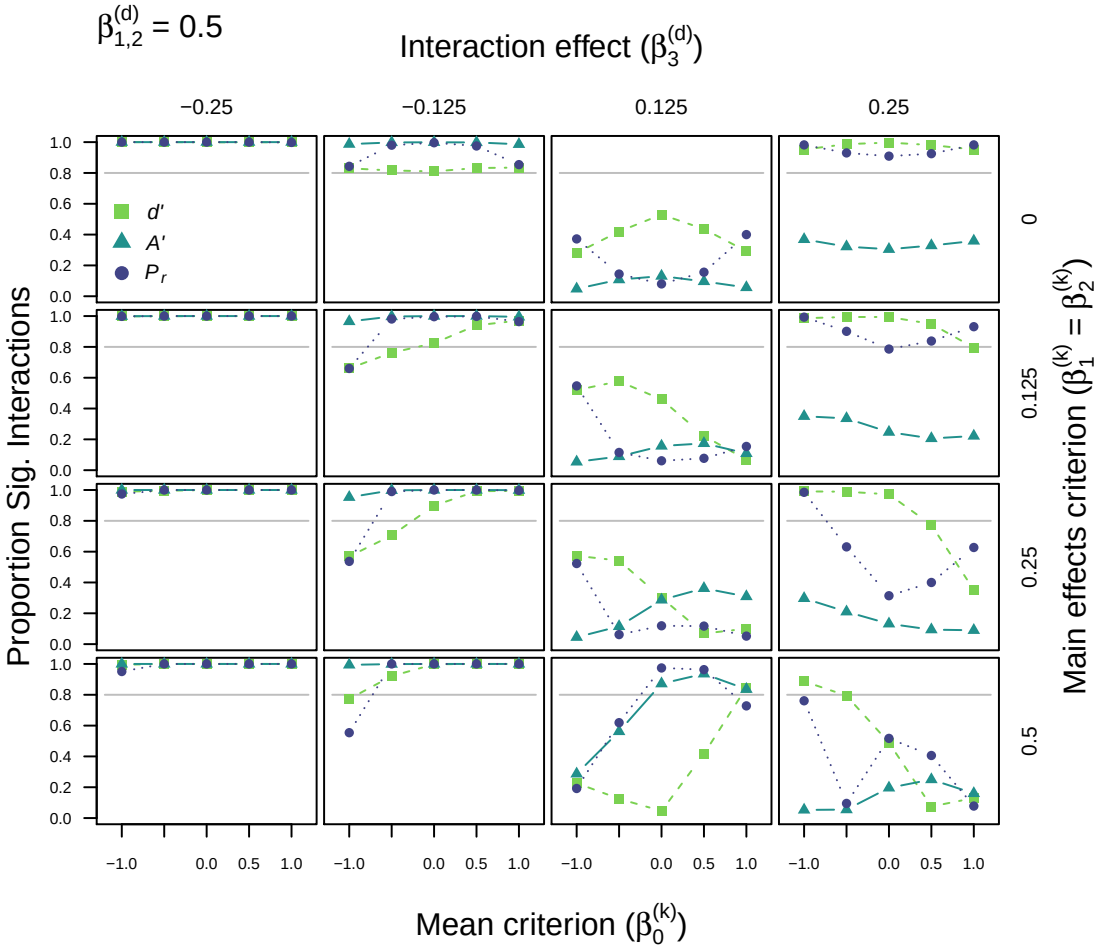

Figure 66: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (GEV-SDT) model. This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and the magnitude of main effects of group and condition on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

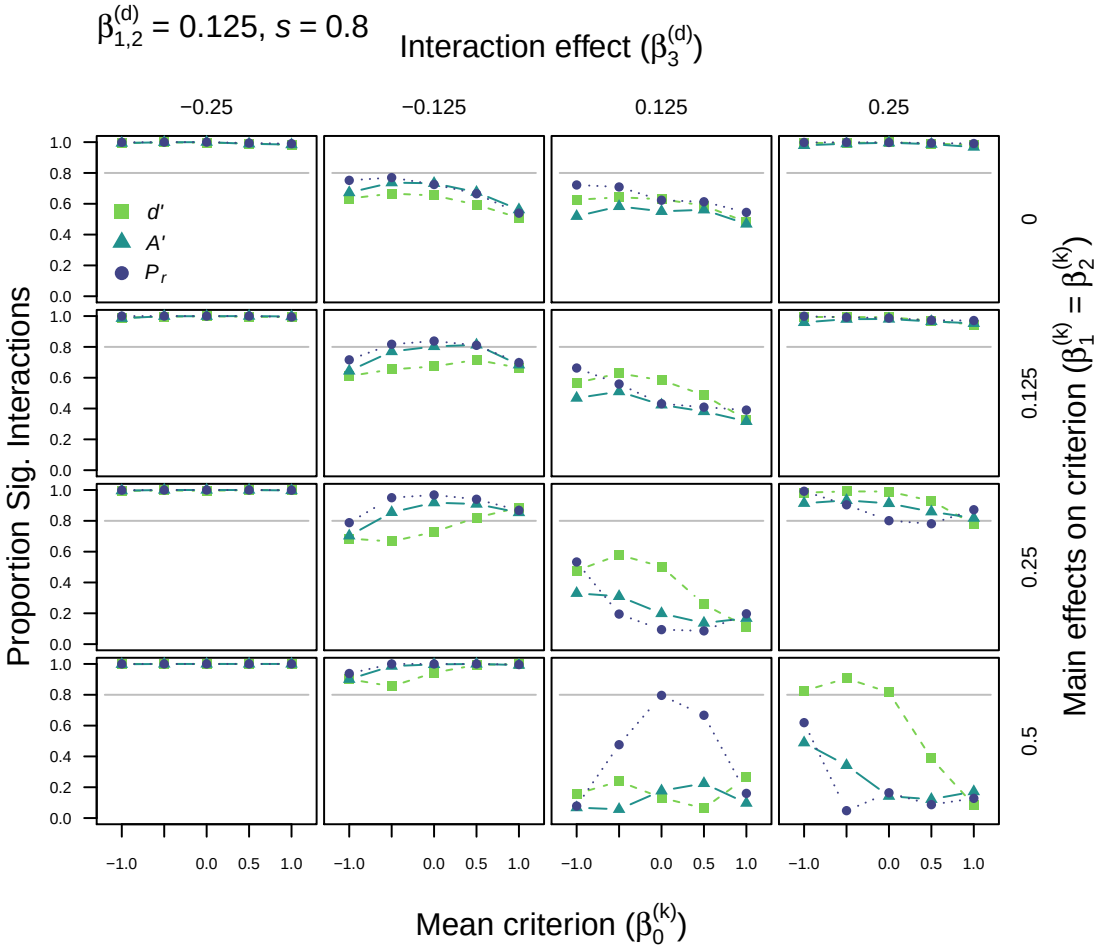

Figure 67: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

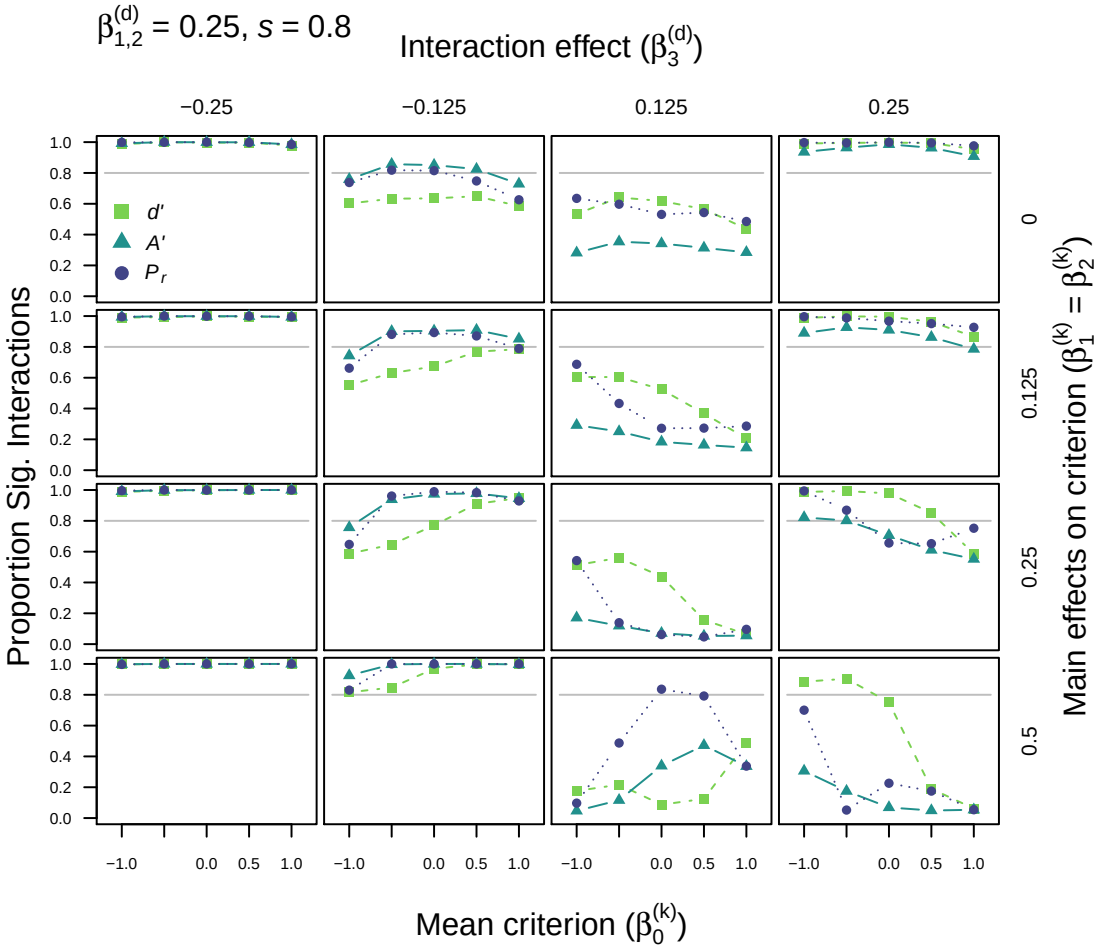

Figure 68: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

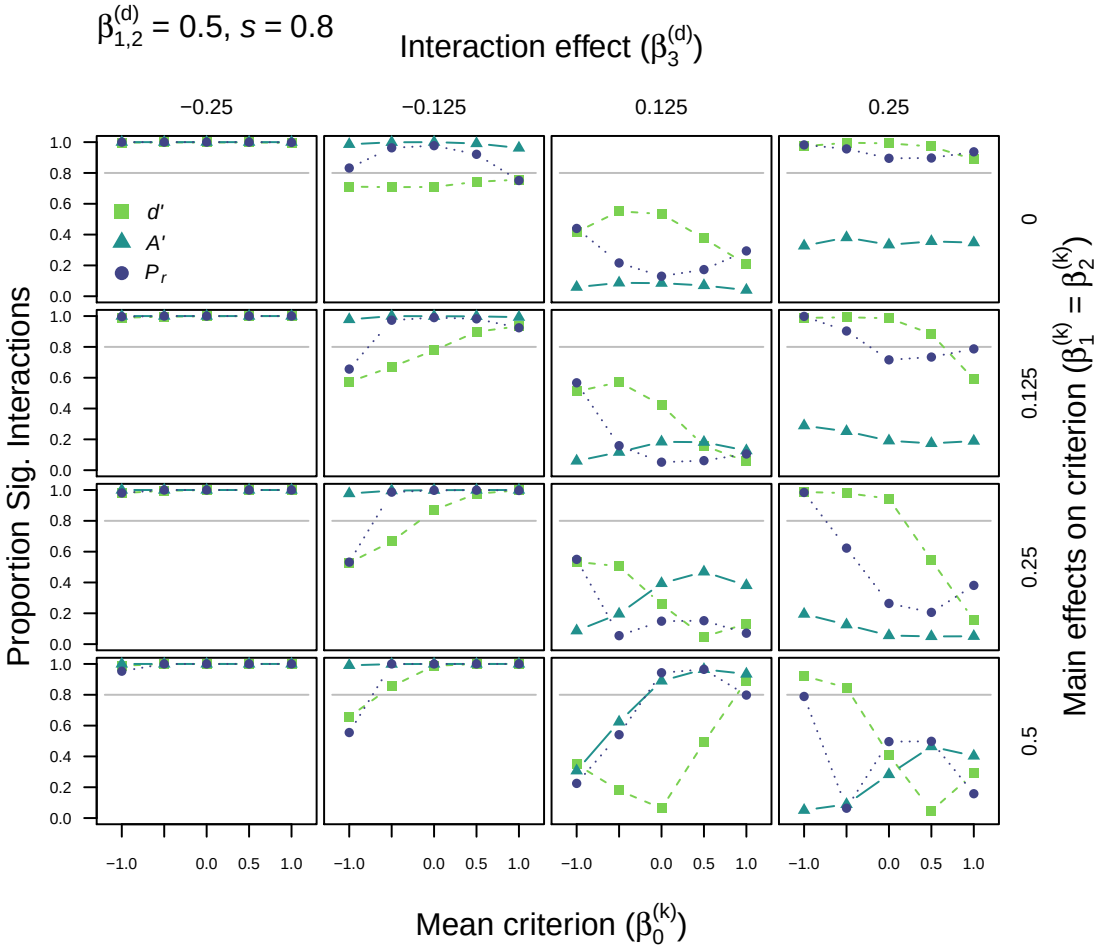

Figure 69: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

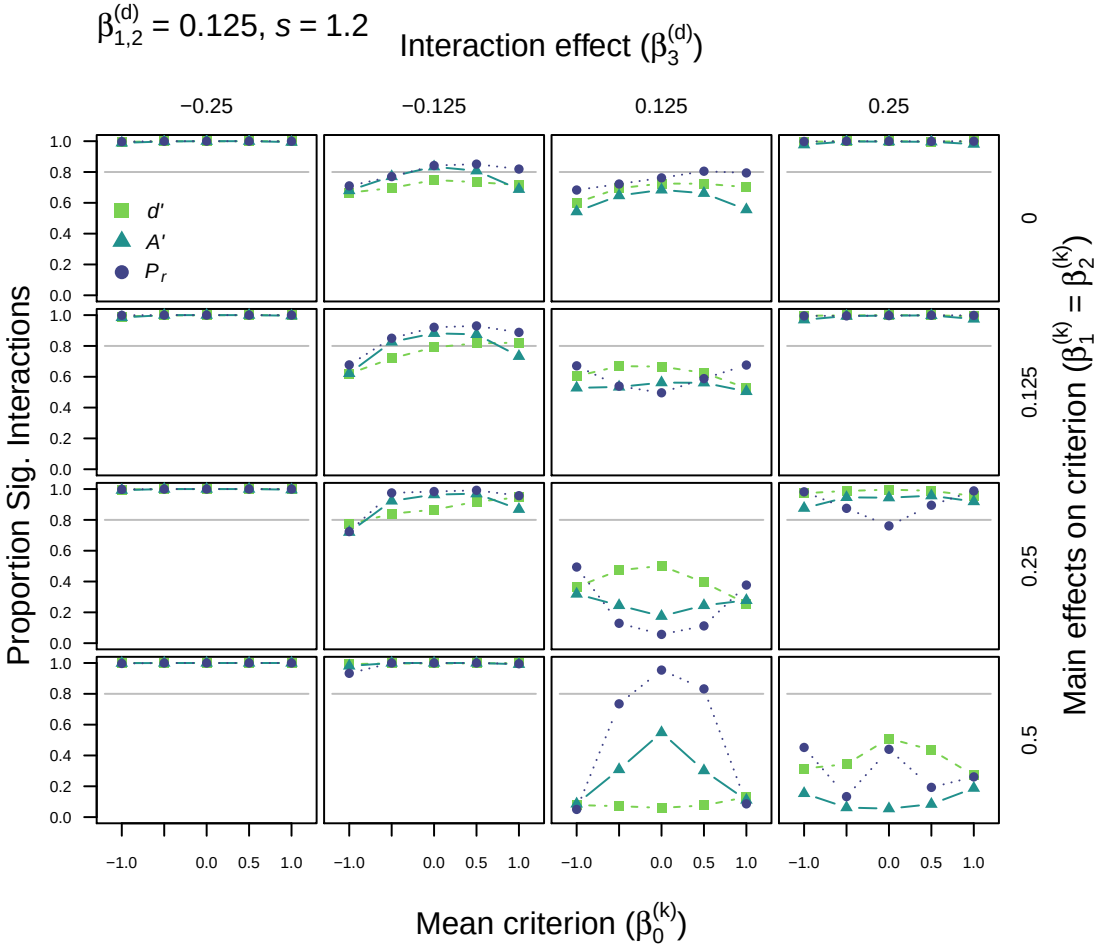

Figure 70: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.125.

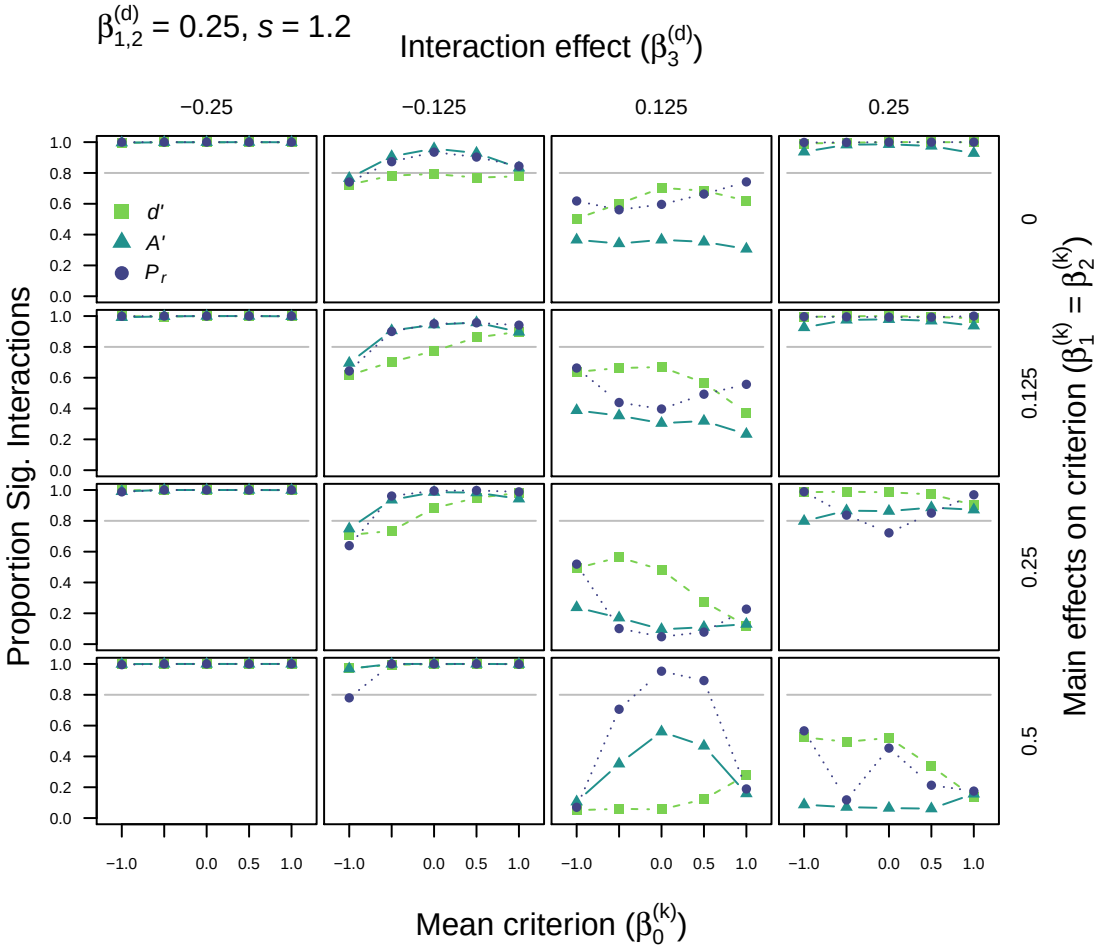

Figure 71: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.25.

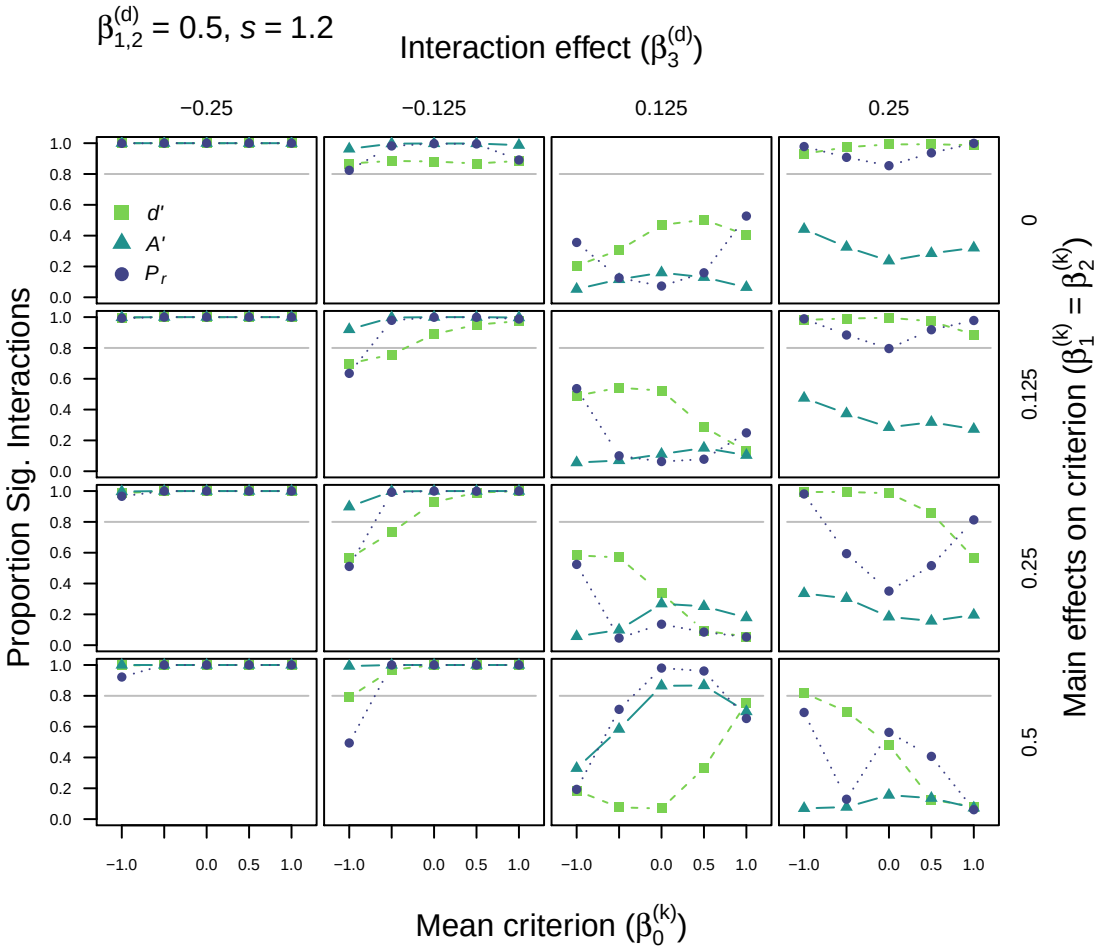

Figure 72: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). Note: this figure depicts the case where main effects on sensitivity equal 0.5.

12 Simulation 12: remove main effects on sensitivity

Finally, simulation 12 was identical to simulation 11, with the exception that there were no underlying main effects on sensitivity. Simulated power in this situation is presented for each generative model in Figures 73 to 76. These figures demonstrate that it is possible to miss interactions purely due to variation in response bias (Rotello et al., 2008; Schooler & Shiffrin, 2005).

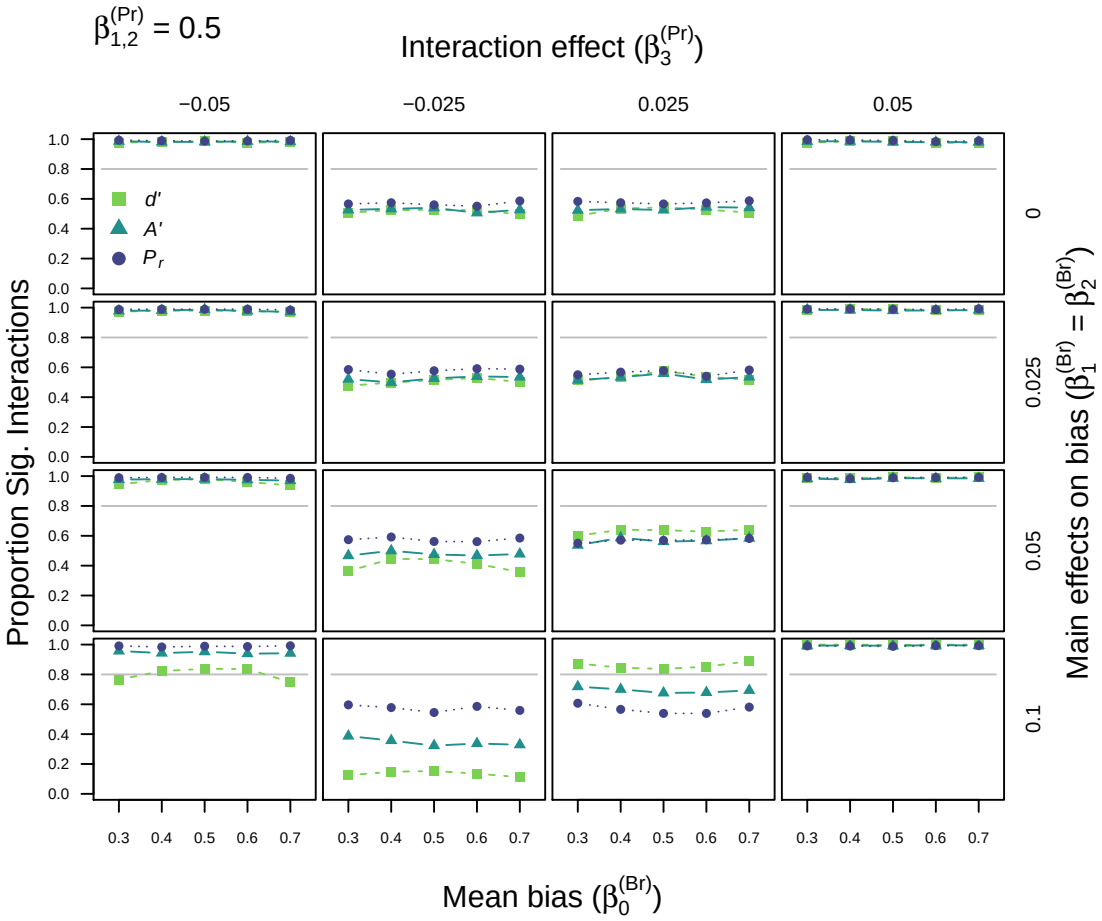

Figure 73: Power for d', A', and P_r with an underlying two-high threshold (THT) model. This simulation varied overall guessing bias (x axis), the magnitude and direction of interaction (left to right) and the magnitude of main effects on bias (top to bottom). There were no main effects on detection probability.

12.1 Two-high threshold

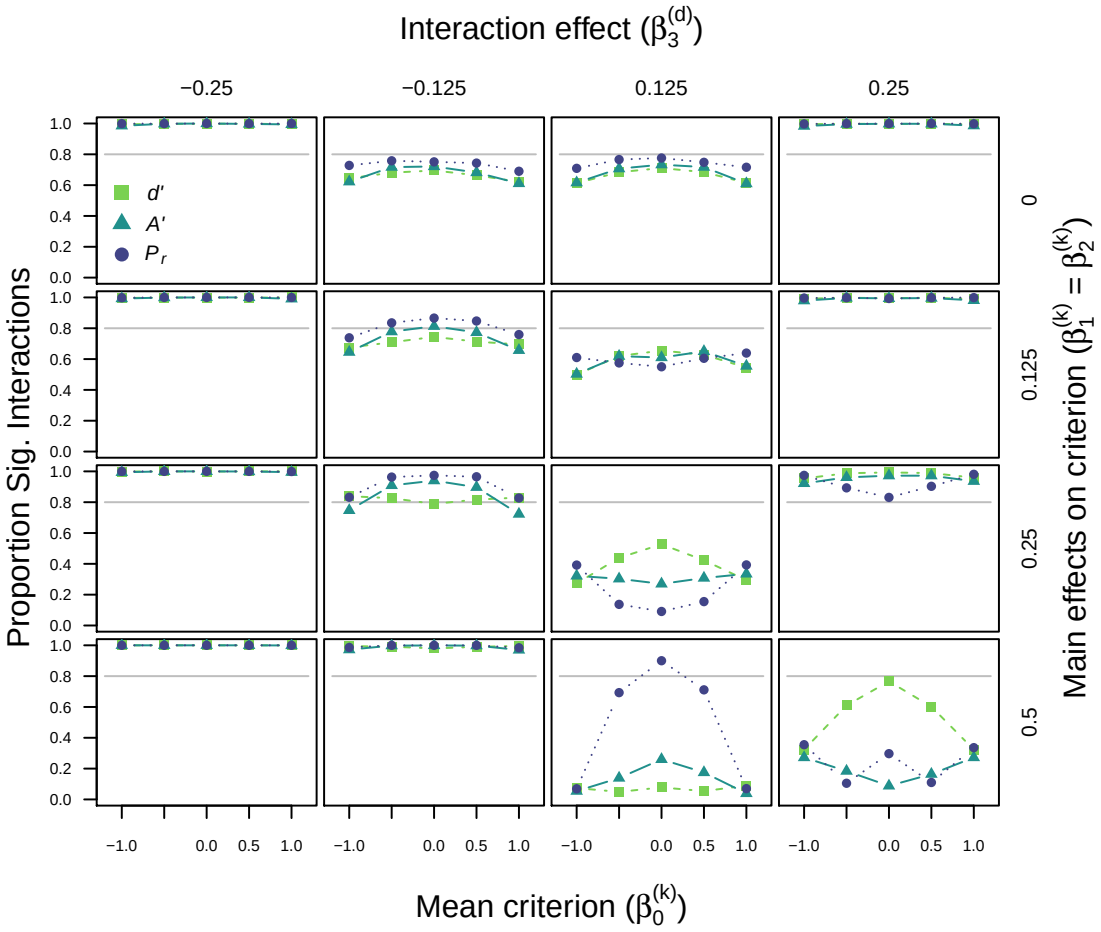

Figure 74: Power for d', A', and P_r with an underlying Gaussian equal variance signal detection theory (SDT) model. This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). There were no main effects on sensitivity.

12.2 Signal detection

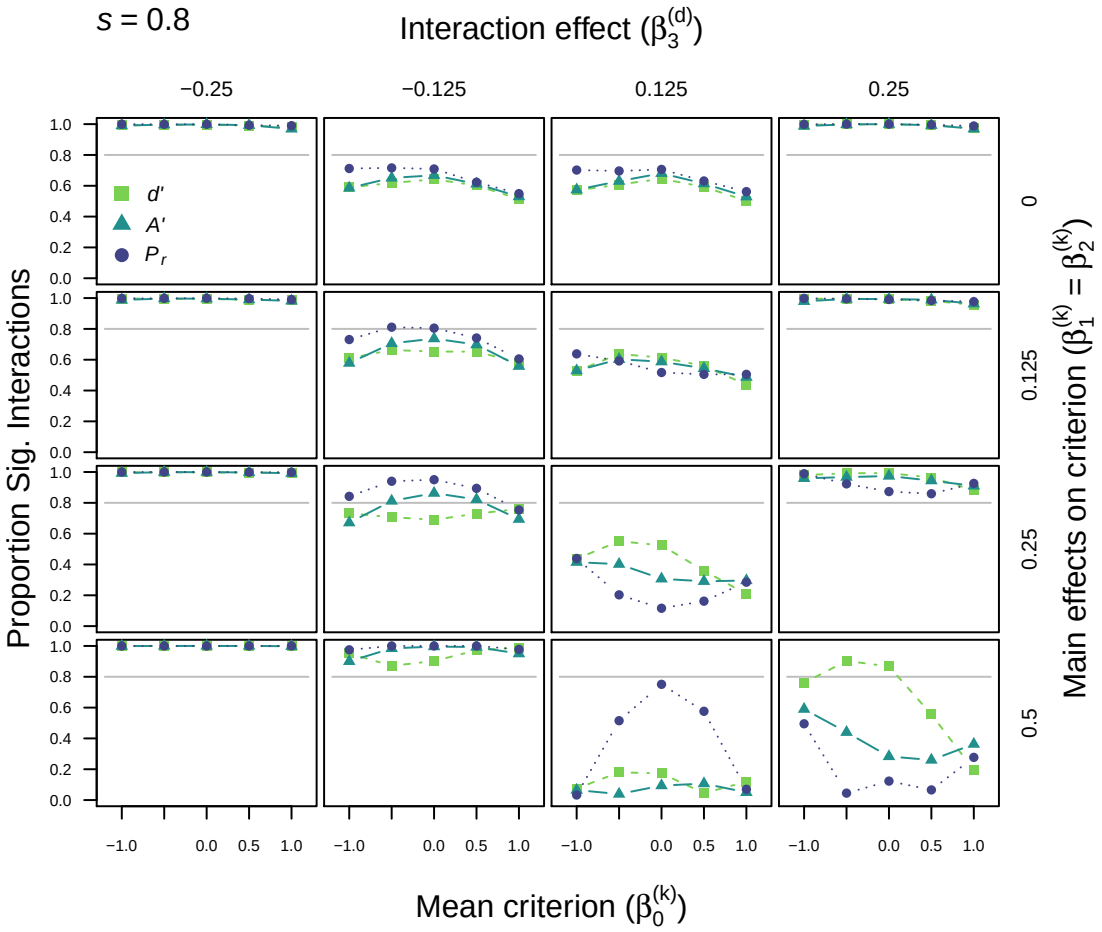

Figure 75: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 0.8 (i.e more variable targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). There were no main effects on sensitivity.

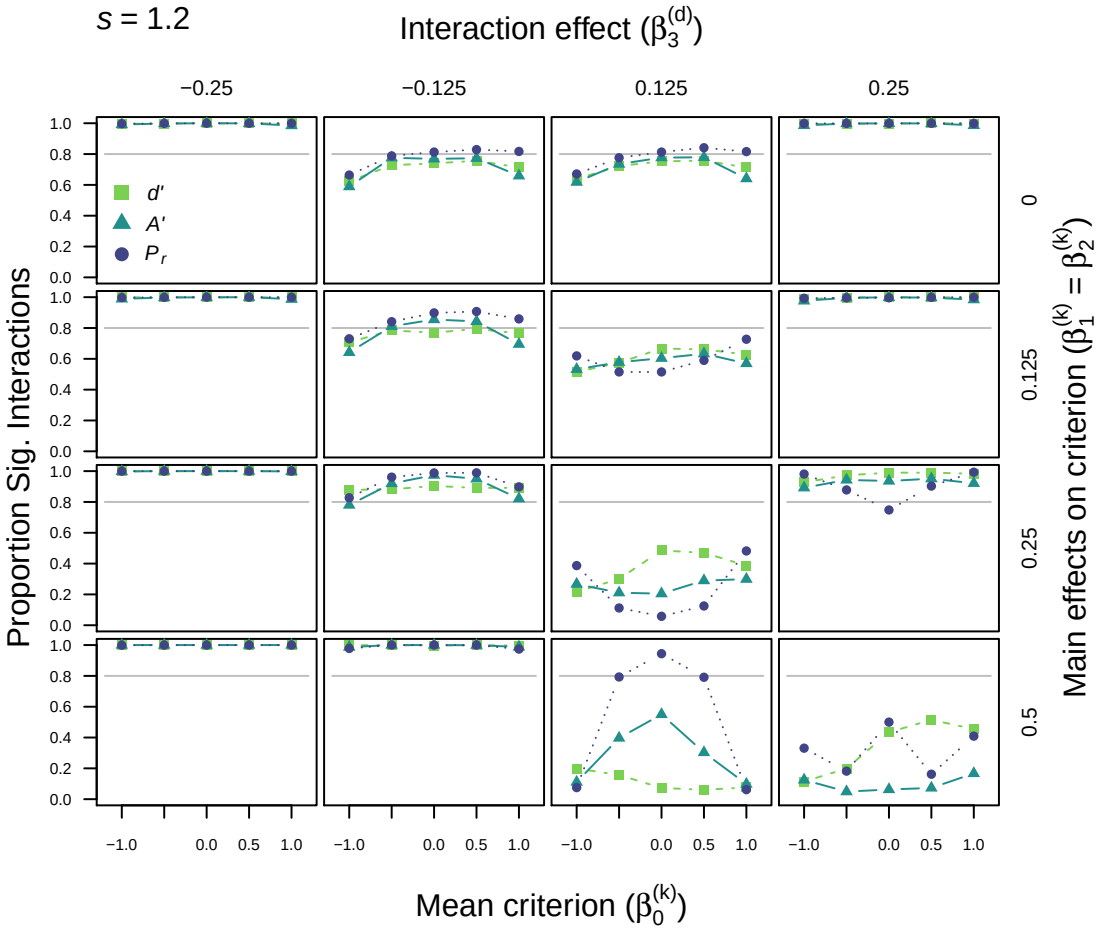

Figure 76: Power for d', A', and P_r with an underlying Gaussian unequal variance signal detection theory (SDT) model. Here s is set to 1.2 (i.e more variable non-targets). This simulation varied overall criterion placement (x axis), the magnitude and direction of interaction (left to right) and main effects on criterion (top to bottom). There were no main effects on sensitivity.

References

Rotello, C. M., Masson, M. E., & Verde, M. F. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics*, *70*(2), 389–401.

Schooler, L. J., & Shiffrin, R. M. (2005). Efficiently measuring recognition performance with sparse data. *Behavior Research Methods*, *37*(1), 3–10.