Results Modeling and Regression Analyses Decision Making With a Goal

Markus Steiner, Nathaniel Phillips

31 1 2018

This is probably the easiest way to keep everyone updated about what we've found so that we can then discuss the story we want to tell. These are the results from the mixed effects model analysis, the check for whether we find the hypothesized pattern in the likelihood of a risky choice given the points earned and the modelling results.

Likelihood of a Risky Choice Given the Points Earned

Let's first visualize the finding and then check with the regression analysis. In order to be able to plot things I've had to create some aggregate measure of the data because on trial level the predictions are binary. So I created ten bins along the range of points a subject had earned and used the mean of the points in that bin as value to plot on the x-axis and the mean of the binary data of risky options chosen to get a likelihood to plot on the y-axis. The finer lines are individual data, the bolder mean lines are fitted with the loess method to depict the data, so it's not an actual model that was fitted and depicted but an aggregatoin over subjects.

Plots

In the goal condition we would expect an increase in the likelihood of a risky choice over points (e.g. house money effect, i.e. the more I've won so far the riskier I can play with it; but note that this is a post hoc explanation), with a drop as soon as the goal is reached. In the no goal there should be no such drop.

Let's first look at the plots over all variance conditions to see the general trend.

Now this looks like overall we find the hypothesized pattern, i.e. we find that in the Goal condition there is a dip in the likelihood of a risky choice when subjects reach (or are very close to reach and probably have enough trials left) the goal. This is optimal, because it minimizes the chance of falling below the goal again. The U-shape of the functions can be explained by good or bad luck, i.e. if they were unlucky they took the risky option and suffered high losses, resulting in negative total points. On the other hand if they were lucky, they could reach a very high total points number by picking the risky option and earning high positive amounts. Because towards the endpoints the data becomes less reliable, the plot window only shows the intervall [-25, 150].

Now let's look at the same plot but separated for the different environments:

The nice thing is that we see the dip around the goal in the goal conditions in all environments, and also the slopes go in the right direction, i.e. in the *Equal* environment the slope in the no goal and in the beginning of the goal condition is pretty close to 0, i.e. they may have learned that the options have similar returns (in fact equal returns) but were risk averse. In the *High* environment they may have learned that the risky option had the higher average return and thus became risky over time. The opposite is true for the *Low* environment.

This is what we see if we look at the same plot but now with trial instead of points on the x-axis.

- We see, at least in some conditions, the hypothesized pattern of the dip in the likelihood of a risky choice at around 100 points. This only occurs in the goal condition. The U-shape is a method artefact.

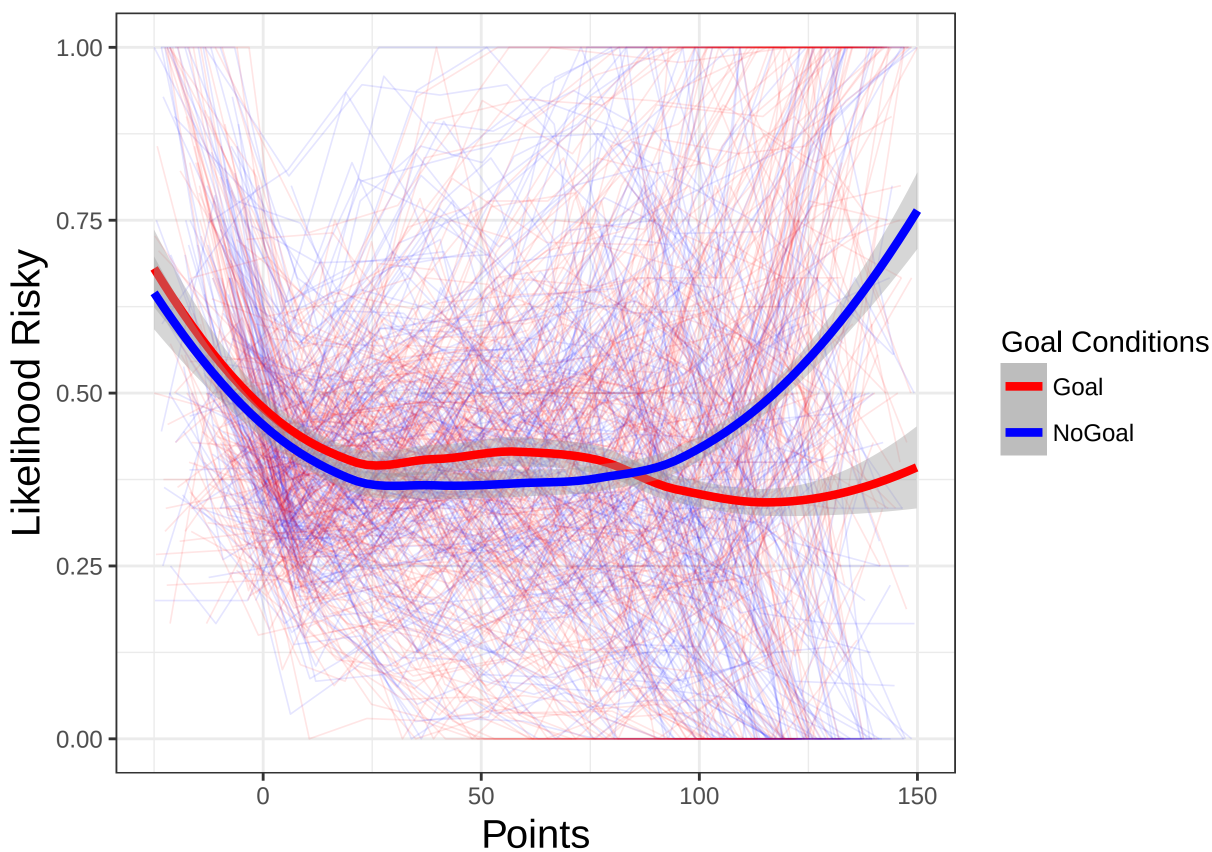


Figure 1: Likelihood of a risky choice by number of points earned, over all conditions.

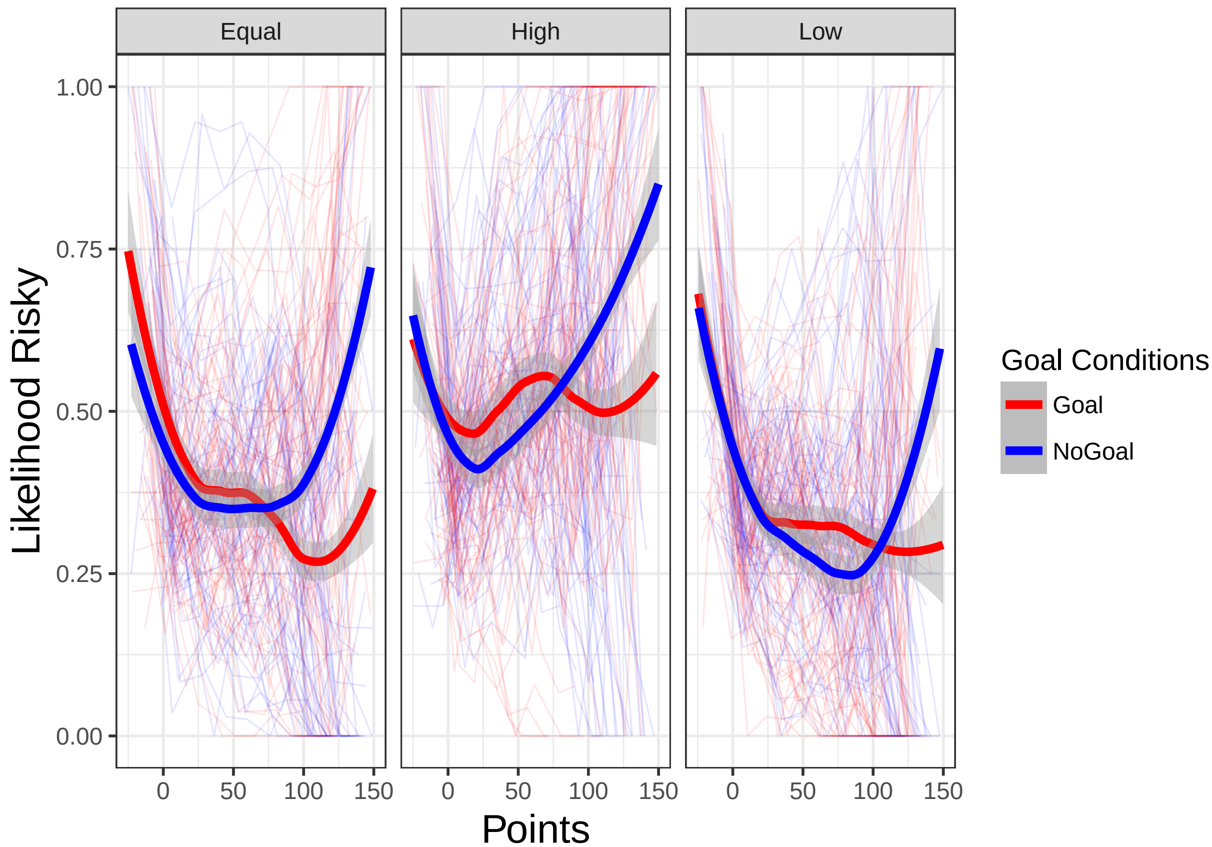


Figure 2: Likelihood of a risky choice by number of points earned, separate environments.

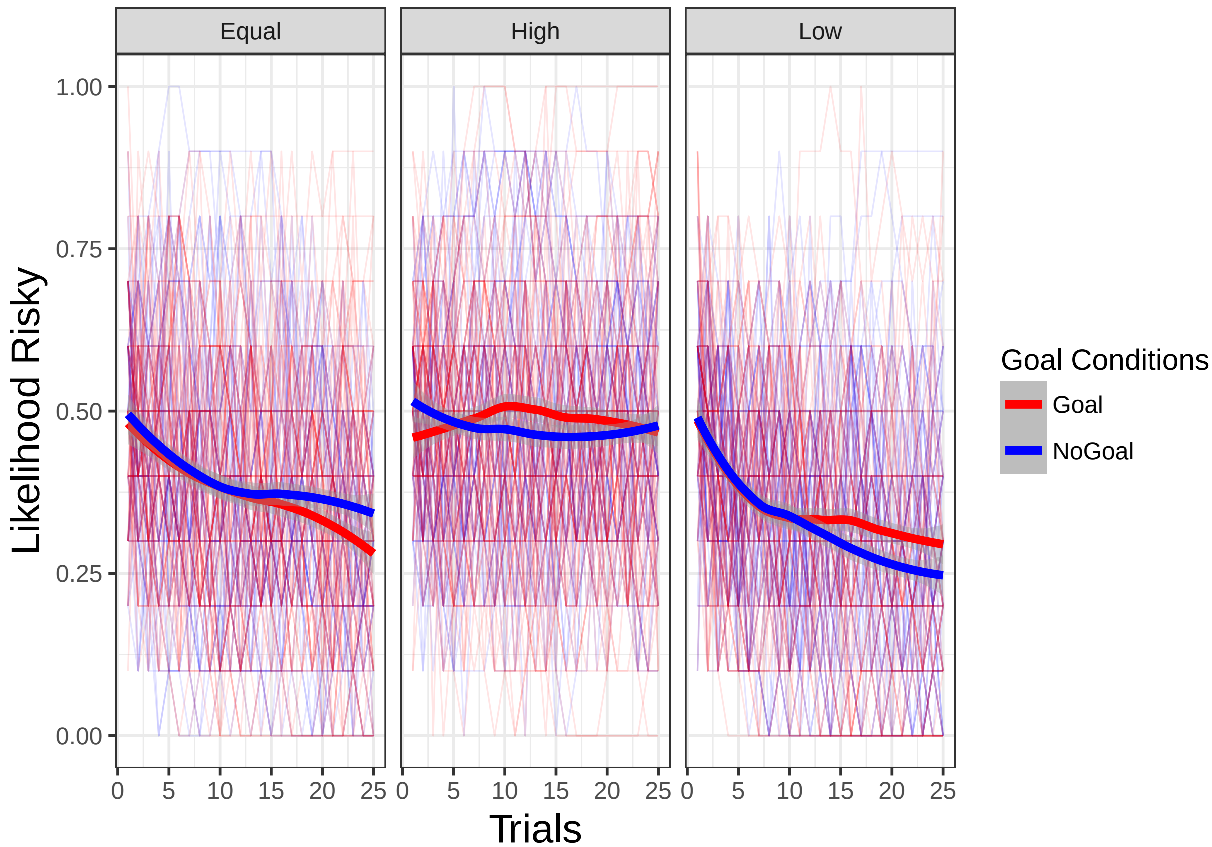


Figure 3: Likelihood of a risky choice by trial number, separated for environments.

Regression Analysis

The dependent variable was whether the risky option was chosen (1) or not (0) at a given trial. Predictors were the goal condition (no goal = 0, goal = 1), whether the goal of 100 points was reached (below goal = 0, above goal = 1), the points earned by a given trial and in the second model, the environment. Random intercepts were included for every person and every game.

We used a generalized mixed effects model. This is how it looked like:

```
# mixed effects model with random intercepts for subjects and games
model <- lme4::glmer(high.var.chosen ~ goal.condition.bin * overGoal.f + points.cum +
                         (1|id.f/game), data = df_trial, family = "binomial")

summary(model)
r2(model)
```

And this is the result of the regression:

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.3923004	0.0665101	-5.898	3.67e-09 ***
goal.condition.bin1	0.0865610	0.0960116	0.902	0.367
overGoal.f1	0.2216200	0.0488954	4.533	5.83e-06 ***
points.cum	-0.0026771	0.0003284	-8.151	3.62e-16 ***
goal.condition.bin1:overGoal.f1	-0.9684303	0.0653238	-14.825	< 2e-16 ***

The coding is $NoGoal = 0$, $Goal = 1$; and belowGoal = 0, overGoal = 1.

It seems that in general, in the *NoGoal* condition, the probability of a risky choice becomes larger, once a subject is above the goal, and decreases slowly with the number of points earned. In the *Goal* condition once the goal is reached, the likelihood of a risky choice decreases. This confirms the patterns shown in the plots. However, we should say that the model was nearly unidentifiable (thus e.g. entering yet another predictor, e.g. the environment may be a problem, but we did it anyway so here's the table).

	Estimate	Std. Error	z value	Pr(>z)
(Intercept)	-0.511628	0.076620	-6.677	2.43e-11 ***
goal.condition.bin1	0.112715	0.077717	1.450	0.1470
overGoal.f1	0.231346	0.043935	5.266	1.40e-07 ***
points.cum	-0.002274	0.000292	-7.788	6.81e-15 ***
variance.conditionHigh	0.488517	0.094269	5.182	2.19e-07 ***
variance.conditionLow	-0.307820	0.095010	-3.240	0.0012 **
goal.condition.bin1:overGoal.f1	-0.929184	0.057912	-16.045	< 2e-16 ***

- The regression analysis showed a significant interaction of overGoal and goal condition, i.e. subjects from the no goal condition were more likely to choose risky after having reached the goal, while subjects from the goal condition where less likely to choose risky after having reached the goal. Likelihood of choosing risky was higher in the *High* environment compared to the *Equal* environment and lower in the *Low* environment compared to the *Equal* environment.

Individual Modeling

We would expect that subjects from the goal condition are best classified by the model that accounts for the goal. The subjects from the no goal model should be best classified by either the mean or the reinforcement learning model.

The next section provides a short overview over the models.

Description of the Models used

Model	Description	Impression parameter	Choice Parameter	Total Free Parameters
Random	Baseline model to compare the others with. Random choice in every trial	-	-	0
NaturalMean	For each option takes the mean of all seen outcomes. Then put's these means into a softmax	-	ϕ	1
RL	Standard reinforcement learning model with one learning parameter for impression updating. Impressions are then put into a softmax.	α (learning rate)	ϕ	2
SampEx_Int_Goal	For each option calculates the mean and standard deviation of the last N (free discrete parameter that was estimated) outcomes, then creates a cummulative (for the number of trials left) normal density and from this derives the probability of sampling a value at least as big as the difference from the current point total to the goal. Then put's these probabilities into a softmax	N (memory capacity)	ϕ	2

Modeling results

Here are tables showing the proportion of people from the Goal and NoGoal condition were classified by which model:

Overall

	Goal	NoGoal
Random	0.180	0.154
NaturalMean	0.148	0.244
RL	0.360	0.421
SampEx_Int_Goal	0.312	0.181

Low Environment

	Goal	NoGoal
Random	0.182	0.072
NaturalMean	0.197	0.275
RL	0.455	0.420
SampEx_Int_Goal	0.167	0.232

High Environment

	Goal	NoGoal
Random	0.197	0.234
NaturalMean	0.115	0.247
RL	0.328	0.364

	Goal	NoGoal
SampEx_Int_Goal	0.361	0.156

Equal Environment

	Goal	NoGoal
Random	0.161	0.147
NaturalMean	0.129	0.213
RL	0.290	0.480
SampEx_Int_Goal	0.419	0.160

The results of two environments (High and Equal) suggest, that the model that accounts for the goal is better in classifying people from the Goal than from the NoGoal condition. Thus also from a modeling perspective we have some evidence that, at least for some subjects, accounting for a person's goal is an important aspect.

Here are some summary tables to show model fits and parameter value statistics.

Over all environments:

model_best	bic_mean	bic_sd	par_Imp_mean	par_Imp_sd	par_Choice_mean	par_Choice_sd	N
NaturalMean	284.6419	57.11349	NA	NA	0.1021640	0.0743124	82
Random	346.5740	0.00000	NA	NA	NA	NA	68
RL	271.3271	60.62504	0.7193733	0.2604189	0.0953607	0.1166225	161
SampEx_Int_Goal	307.5428	30.28930	2.8787879	3.6902318	1.7736558	0.7426561	99

Separated for the three environments

model_best	variance.condition	bic_mean	bic_sd	par_Imp_mean	par_Imp_sd	par_Choice_mean	par_Choice_sd	N
NaturalMean	Equal	288.9867	54.12785	NA	NA	0.0951604	0.0701292	24
NaturalMean	High	289.3647	51.91944	NA	NA	0.1104961	0.0897531	26
NaturalMean	Low	277.5460	63.95490	NA	NA	0.1006469	0.0646944	32
Random	Equal	346.5740	0.00000	NA	NA	NA	NA	21
Random	High	346.5740	0.00000	NA	NA	NA	NA	30
Random	Low	346.5740	0.00000	NA	NA	NA	NA	17
RL	Equal	276.3863	54.87985	0.6979056	0.2566165	0.0922045	0.1094558	54
RL	High	280.6715	48.69127	0.7632377	0.2534000	0.0737897	0.0605315	48
RL	Low	259.0944	72.23123	0.7033355	0.2694299	0.1157988	0.1509413	59
SampEx_Int_Goal	Equal	311.3188	29.13052	3.1842105	3.8963272	1.6562858	0.8051688	38
SampEx_Int_Goal	High	305.8229	29.23471	2.6176471	3.4816811	1.9621618	0.7205251	34
SampEx_Int_Goal	Low	304.3941	33.62546	2.7777778	3.7553381	1.7014653	0.6521926	27

- We saw that subjects from the goal condition are more often (usually around twice as often) best classified by the goal model, however the majority in both groups is usually best classified by standard reinforcement learnin model. The good thing is that in the goal condition the proportion best classified by the goal model is substantial (around one third) providing evidence that it can be important to incorporate a subject's goal in a model. Mean parameter values didn't differ much over environments. The memory capacity parameter in the SampEx model was on average between 2.5 and 3 samples.

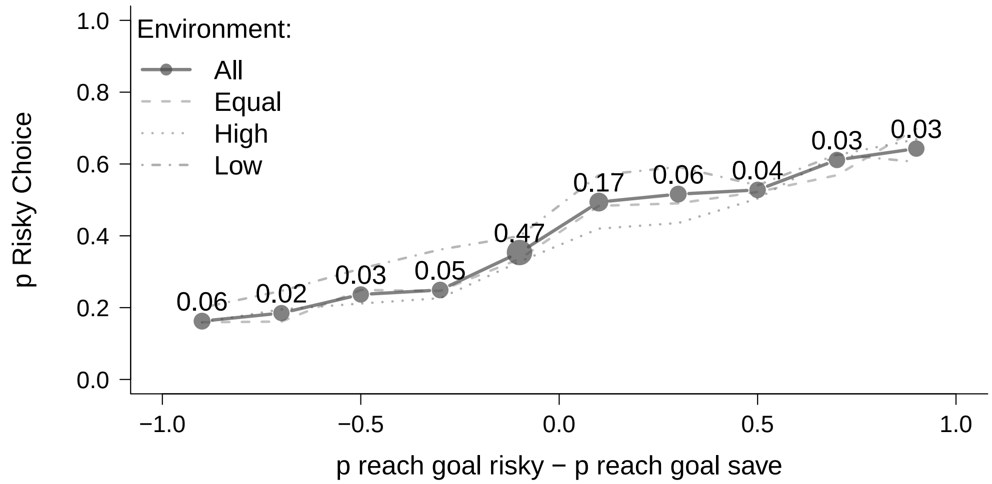
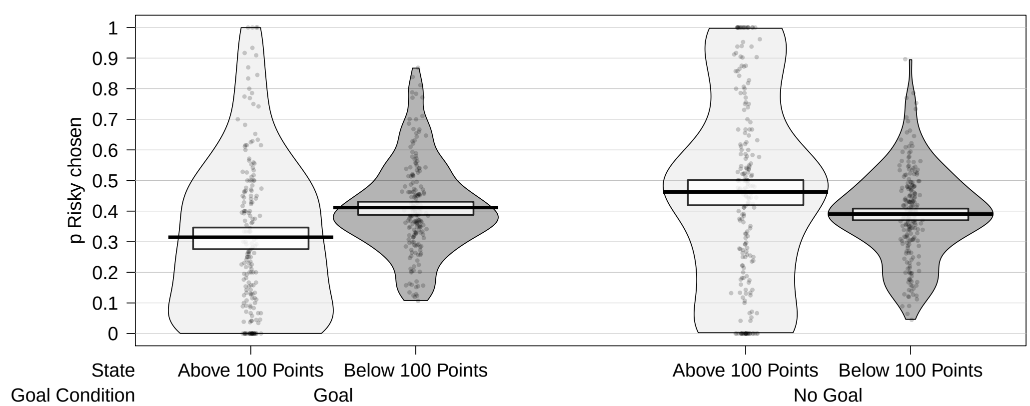


Figure 4: Probability of choosing the risky option, given the RST evidence strengths for that option. Only data from goal condition was used. Point sizes and numbers above the points indicate the proportion of occurrences within these bins of evidence strengths.

Previous/ Preregistered Analysis Results

Here are the Plots from the previous analyses that are also interesting and could be used here.

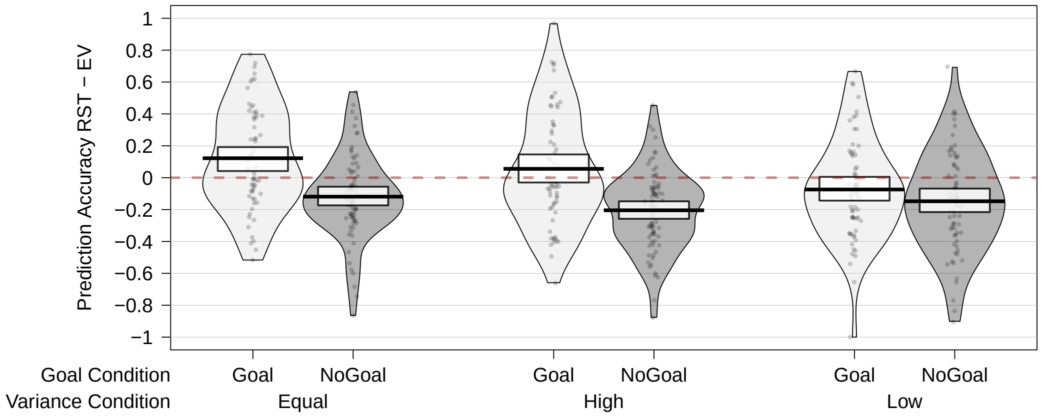
```
include_graphics("/Users/msteiner/Dropbox/Masterarbeit/eegoals/plot/pRiskyAboveUnderGoalNoVarCond.pdf")
```



\begin{figure}
\caption{Proportion of risky options chosen, separated for goal condition and the state. 100 Points corresponded to the goal in the goal condition. The effect was roughly the same over environments. Bold horizontal lines indicate the mean. Boxes indicate 95% bayesian highest density intervals.} \end{figure}

```
include_graphics("/Users/msteiner/Dropbox/Masterarbeit/eegoals/plot/EvidenceStrengthsRisky.pdf")
```

```
include_graphics("/Users/msteiner/Dropbox/Masterarbeit/eegoals/plot/predictionAccuraciesDiffRST-EV.pdf")
```



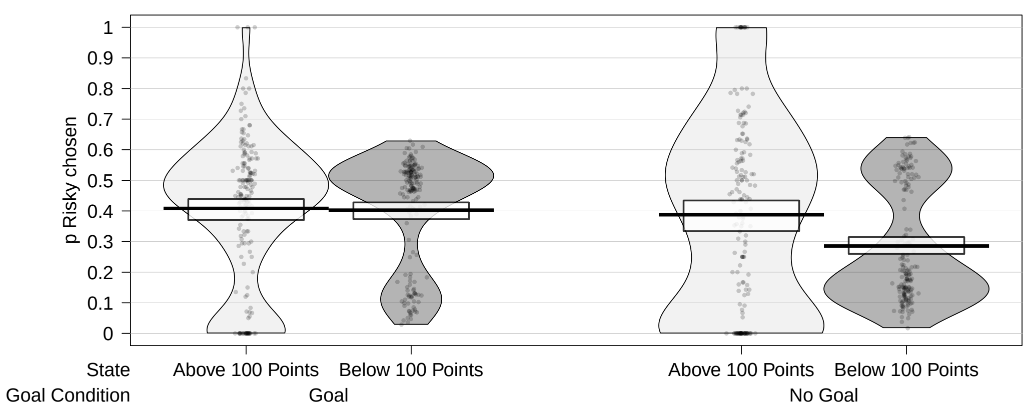
\begin{figure}
\caption{Differences in prediction accuracy rates ($pA_{RST} - pA_{EV}$) in trials in which the model prediction differed, separated for both goals and environments. A value above 0 (above the dashed line), indicates higher prediction accuracy rate of RST, a value below 0 a higher prediction accuracy rate of EV. Bold horizontal lines indicate the mean. Boxes indicate 95% Bayesian highest density intervals.} \end{figure}

Simulation based plots for the same analyses

First the plots for the likelihood of a risky choice given the number of points:

And now the same but with trial on the x-axis

```
include_graphics("/Users/msteiner/Dropbox/Masterarbeit/eegoals/plot/pRiskyAboveUnderGoalNoVarCondSimula
```



\begin{figure}
\caption{Plot of simulated data. Proportion of risky options chosen, separated for goal condition and the state. 100 Points corresponded to the goal in the goal condition. The effect was roughly the same over environments. Bold horizontal lines indicate the mean. Boxes indicate 95% bayesian highest density intervals.} \end{figure}

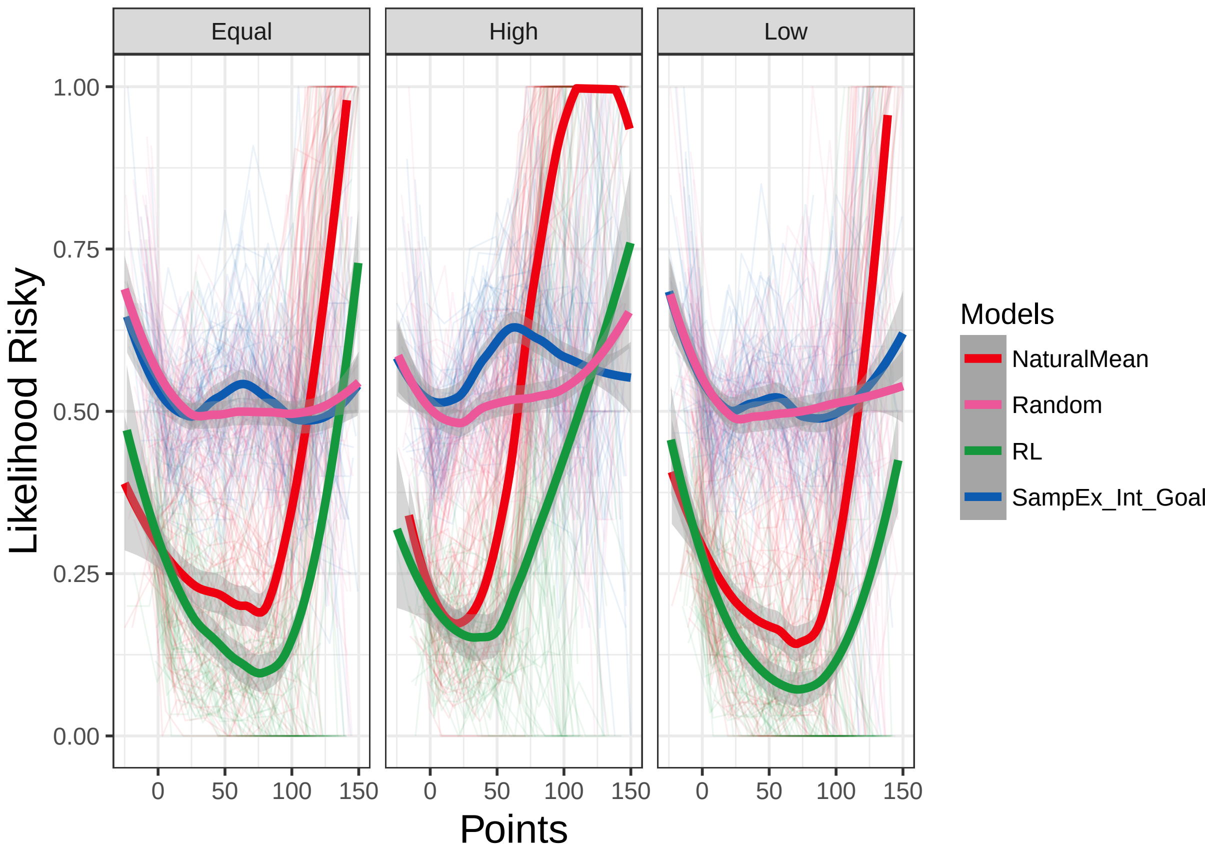


Figure 5: Simulation Data: Likelihood of a risky choice by number of points earned, separated for environments.

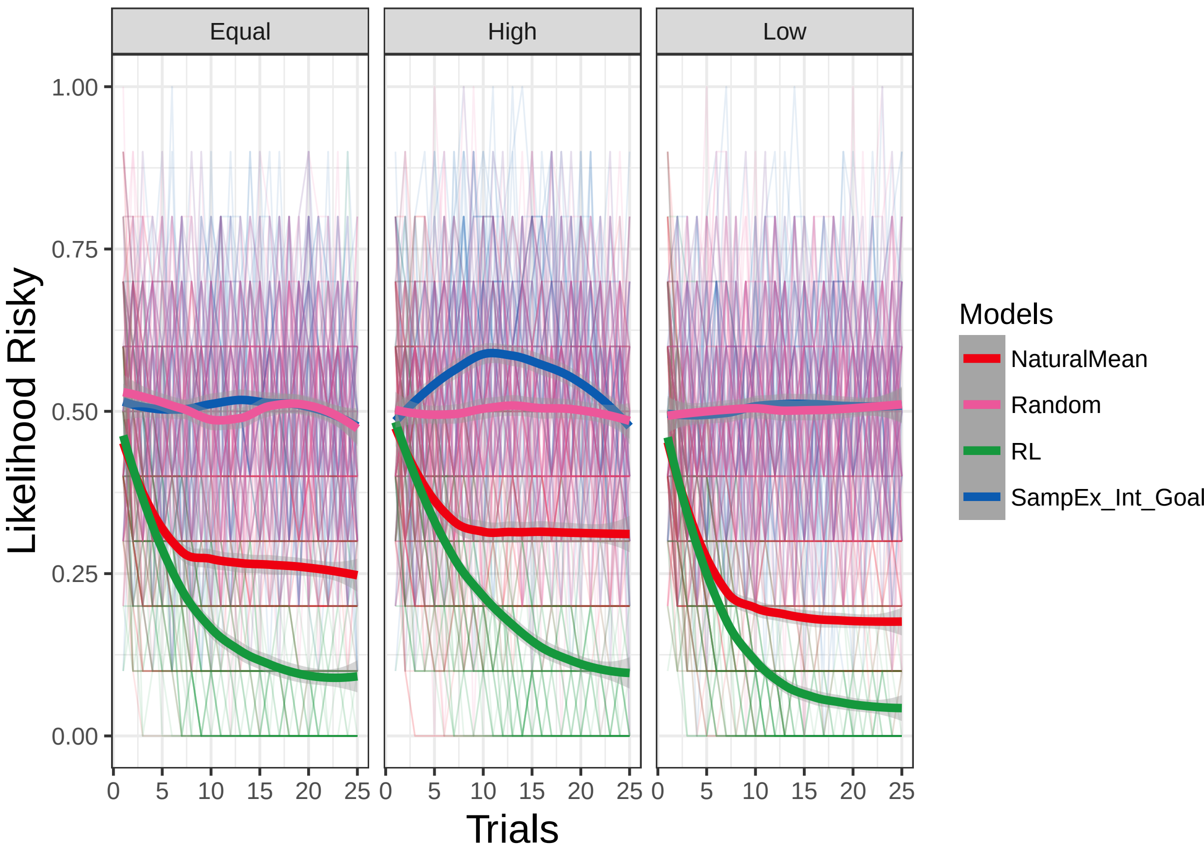


Figure 6: Simulation Data: Likelihood of a risky choice by trial number, separated for environments.