We wish to represent the joint probability distribution over actions a^i and reaction times rt^i given the state s^i for trial i in the task $p(a^i, rt^i|s^i)$
$p(a^i, rt^i|s^i) = \frac{p(a^i,rt^i,s^i)}{p(s^i)}$ by the chain rule for probabilities
$= \frac{p(rt^i|a^i,s^i)p(a^i|s^i)p(s^i)}{p(s^i)}$ by the chain rule for probabilities
$= p(rt^i|a^i, s^i)p(a^i|s^i)$
$= p(rt^i|a^i = 0, s^i)p(a^i = 0|s^i) + p(rt^i|a^i = 1, s^i)p(a^i = 1|s^i)$ where $a^i = 0$ corresponds to withholding a response and $a^i = 1$ corresponds to a response

Let's code $rt^i = NaN$ when $a^i = 0$. Then

$$p(a^i, rt^i|s^i) = \begin{cases} p(rt^i|a^i = 1, s^i)p(a^i = 1|s^i), & \text{if } rt \neq NaN \\ p(rt^i|a^i = 0, s^i)p(a^i = 0|s^i), & \text{if } rt = NaN \end{cases}$$

Because $p(rt^i = NaN|a^i = 0, s^i) = 1$, we can simplify

$$p(a^i, rt^i|s^i) = \begin{cases} p(rt^i|s^i)p(a^i = 1|s^i), & \text{if } rt \neq NaN \\ p(a^i = 0|s^i), & \text{if } rt = NaN \end{cases}$$

Next we consider each term separately. We can represent $p(a^i = a|s^i)$ for $a \in 0, 1$ using the softmax rule
$p(a^i = a|s^i) = \frac{exp(Q(s^i,a)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}$

Next, if we assumed reaction times are linearly related to the Q value for the go action $rt^i = \beta * Q(s^i, a^i) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$
This implies that
$p(rt^i|Q(s^i, a^i)) = \frac{1}{2\pi\sigma} exp(\frac{-(rt^i - Q(s^i,a^i)*\beta)^2}{2\sigma^2})$

Note that this equation generalizes to multiple linear regression and simply says that the probability of an observation is specified by a Gaussian PDF centered on the regression line. Note that the minimum of this function does not depend on the variance of the gaussian distribution, so this formulation is equivalent to the standard least squared minimization approach. However, we will need to use the variance when computing the joint probability distribution over actions and rts.

We wish to maximize the log likelihood of the parameters θ.
$log(L(\theta)) = \sum log(p(a^i, rt^i|s^i; \theta)) =$

$$\begin{cases} \sum log(p(rt^i|s^i)) + \sum log(p(a^i = 1|s^i)), & \text{if } rt \neq NaN \\ \sum log(p(a^i = 0|s^i)), & \text{if } rt = NaN \end{cases}$$

We can substitute in the softmax and the probabilistic regression equation from above to get

$$\begin{cases} \sum log(\frac{1}{2\pi\sigma} exp(\frac{-(rt^i - Q(s^i,a^i)*\beta)^2}{2\sigma^2})) + \sum log(\frac{exp(Q(s^i,a=1)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt \neq NaN \\ \sum log(\frac{exp(Q(s^i,a=0)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt = NaN \end{cases}$$

If there m trials, then this simplifies to:

$$\begin{cases} mlog(\frac{1}{2\pi\sigma}) - \frac{1}{2\sigma^2} \sum (rt^i - Q(s^i, a^i) * \beta)^2 + \sum log(\frac{exp(Q(s^i,a=1)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt \neq NaN \\ \sum log(\frac{exp(Q(s^i,a=0)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt = NaN \end{cases}$$

and since we wish maximize $L(\theta)$, then we can drop the first term because it does not depend on θ

$$\begin{cases} -\frac{1}{2\sigma^2} \sum (rt^i - Q(s^i, a^i) * \beta)^2 + \sum log(\frac{exp(Q(s^i,a=1)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt \neq NaN \\ \sum log(\frac{exp(Q(s^i,a=0)\tau)}{\sum_{a\prime} exp(Q(s^i,a\prime)\tau)}), & \text{if } rt = NaN \end{cases}$$

At each iteration of model fitting, σ can be computed exactly as the variance of the residual RT distribution. Ultimately, maximizing the likelihood function for the joint distribution on actions and reaction times reduces to minimizing the weighted sum of their independent loss functions:

$max(LL(actions, RTs|state; \theta)) = min(\frac{1}{2\sigma^2} Loss(Regression) + Loss(QLearning))$