



this doesn't work, it doesn't properly equalize
 if someone is wrong, if they input happy
 text into standard, they are guaranteed
 to be above 1, so always some crazy increase

baseline model = $f(x)$, standard model = $g(x)$

previous text = x_1

now text = x_2

case 1: if $f(x_2) > 0.5$,

if $f(x_1) > 0.5$

$g(x_2) - g(x_1)$

elif $f(x_1) < 0.5$

else $(g(x_2) + 1) - f(x_1)$

if $f(x_1) > 0.5$

$f(x_2) - (g(x_1) + 1)$

positive number

else

$f(x_2) - f(x_1)$