Supplemental Material

Performance and Response Times

Subjects' predictions improved throughout the training phase (Fig S1). The population correlation between absolute error and trial was -0.244 (p = 0.014), indicating that absolute errors decreased as trial number increased. In the test phase, however, subjects did not improve any further (r = -0.009, p = 0.926).

The absence of any improvement in the test phase raised the question of whether the explicit feedback provided to subjects was doing more harm than good. Would subjects have continued improving if the extra information was not there? In order to test this, we ran an alternative experiment (N = 33) with an identical training phase but a test phase without the onscreen feedback or eye-tracking. Performance in this alternative training phase was similar to the original training phase, with marked improvement over 100 trials (correlation between average absolute error across subjects and trial, r = -0.285, p = 0.004). However, in the alternative test phase, subjects displayed a slight decline in performance (correlation between average absolute error and trial, r = 0.166, p = 0.099; Fig S1).

Additionally, for each phase, we regressed absolute error on the trial number, interacted with the condition (original or alternative) with clustered standard errors at the subject level (the mixed-effects model did not converge). For both phases, there is no significant difference between the effect of trial on absolute error between the original and alternative versions (*training*: β = -0.0008, p = 0.59; *test*: β = 0.002, p = 0.18). If anything, as suggested by the simple correlations, subjects in the alternative experiment performed (slightly) worse than those in the original experiment over time in the test phase. Therefore, we have evidence to suggest that the extra information was not harmful/distracting to subjects during their estimates, and may

have in fact been somewhat beneficial. All of the remaining analyses will only include the original data.

Subjects varied considerably in their performance in the training (average absolute error range: [1.73, 2.68], M = 2.19, SD = 0.23) and test (average absolute error range: [1.59, 2.67], M = 2.08, SD = 0.27) phases. Subjects also varied in the time they took to make their judgments (*training*: average response time [RT] range: [2.23, 9.73], M = 3.81, SD = 1.27; *test*: average response time [RT] range: [2.87, 18.72], M = 4.66, SD = 2.98). In the training phase, all but one subject took less than 6 seconds, on average, to make their judgments. Excluding the one subject who took substantially longer than average (4.7 standard deviations above the mean; M = 9.73, SD = 4.83), there was a significant negative correlation (r = -0.47, p = 0.005) between average absolute error and average response time (RT, Fig 3a). That is, subjects who took more time to respond performed better (on average). However, this relationship does not hold within a subject (i.e. most subjects did not perform better on trials with longer RTs, Fig 3c) and (surprisingly) does not extend to the test phase (with outlier: r = 0.04, p = 0.82; without outlier: r = -0.20, p = 0.27, Fig 3b). The lack of correlation at the individual level is not surprising since there is a counteracting force of trial-level difficulty, which produces a positive correlation between RT and error (Krajbich, Hare, Bartling, Morisihima, & Fehr, 2015). The interaction of these two counteracting forces is a potential explanation for the slight U-shape observed in Fig 3d. Detrimental "overthinking" could also explain this trend (Gill & Prowse, 2017; Moritz, Siemsen, & Kremer, 2013).

The significant across-subject relationship in the training phase suggests that there might be a systematic difference between subjects in the strategies they use. It is possible, for instance, that the subjects who took longer (on average) devoted more attention to the task and were

therefore able to better learn the passive subject's preferences, as in the standard speed-accuracy tradeoff (Wickelgren, 1977).

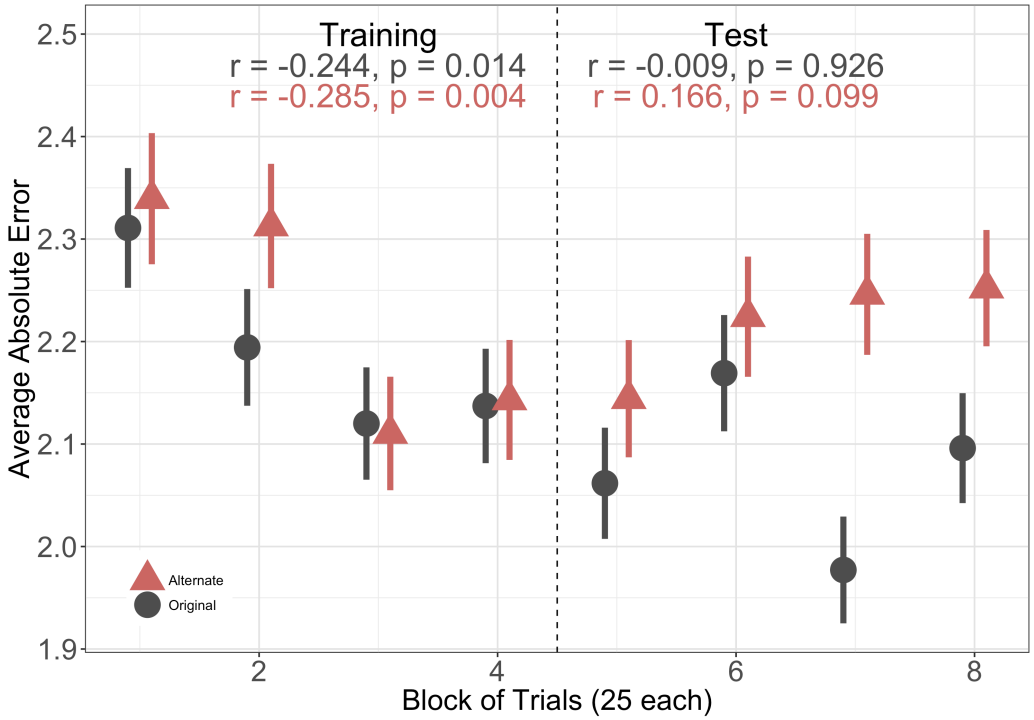

Figure S1. Subject performance. Subjects improved over time in the training phase, but their performance was constant in the test phase. However, results from an alternative version of the experiment (without training phase feedback during the test phase) suggests this was likely due to a plateau effect and not the presence of training phase films onscreen. The black triangles are from the original experiment and the red circles are from the alternative version (error bars are s.e.m. across subjects). Correlations are computed across all data.

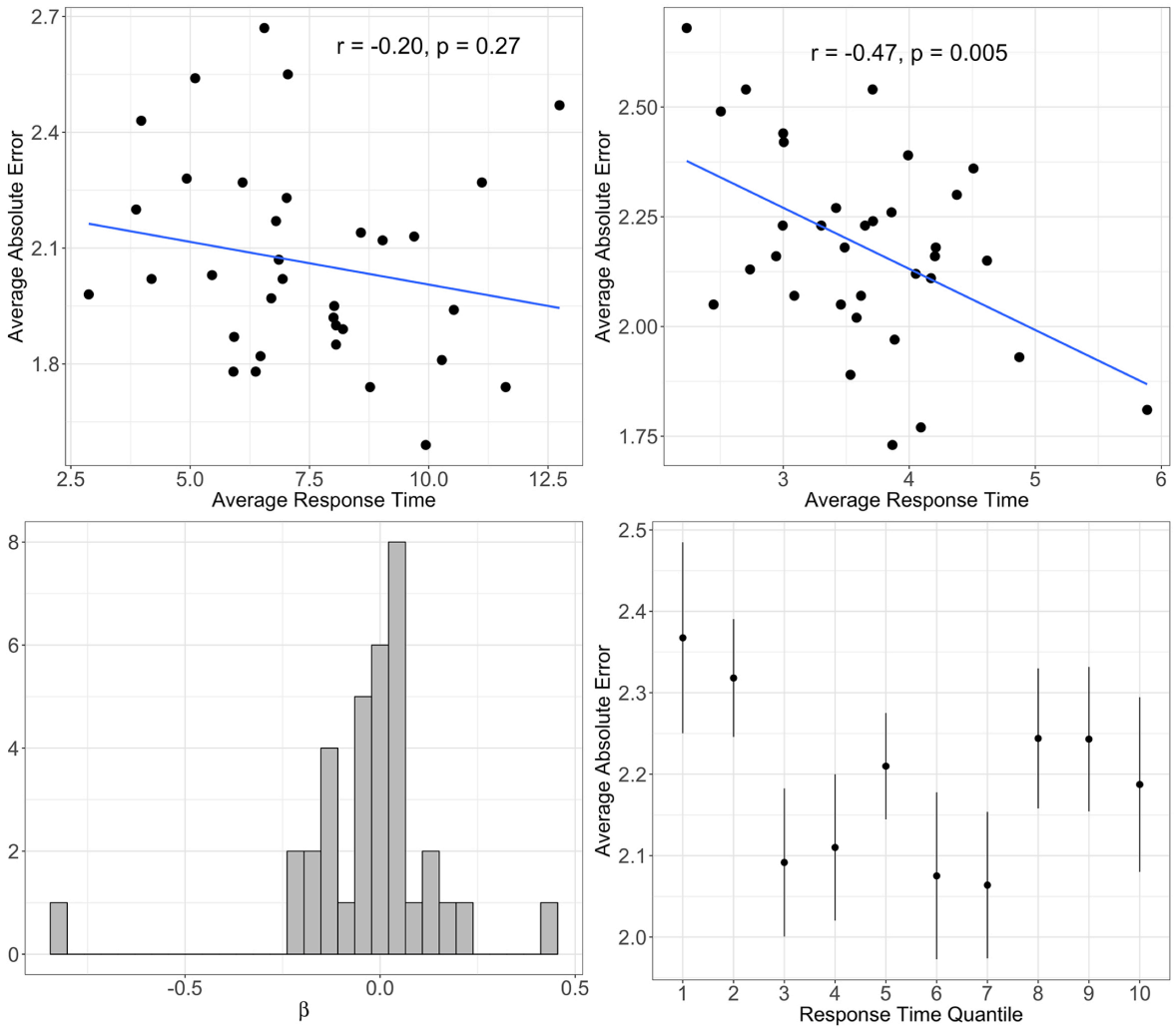

Figure S2. Subject-level performance in the training and test phases as a function of response time. (a) Subjects who took longer, on average, during the training phase tended to perform better on the task (r = -0.47, p = 0.005), (b) but this did not extend to the test phase (r = -0.2, p = 0.27). Each point represents a subject and the blue line is a simple fitted regression through the points. Moreover, this relationship does not typically hold within a subject (c), as evidenced by the abundance of near-zero and positive linear coefficients in individual-level regressions of AbsoluteError on ResponseTime (M = -0.03, SD = 0.19). (d) When we separate the RTs at the subject level into quantiles (at the 10^{th}, 20^{th}, 30^{th}, etc. percentiles), we see a slight U-shape in the average absolute error. Error bars are s.e.m. across subjects.

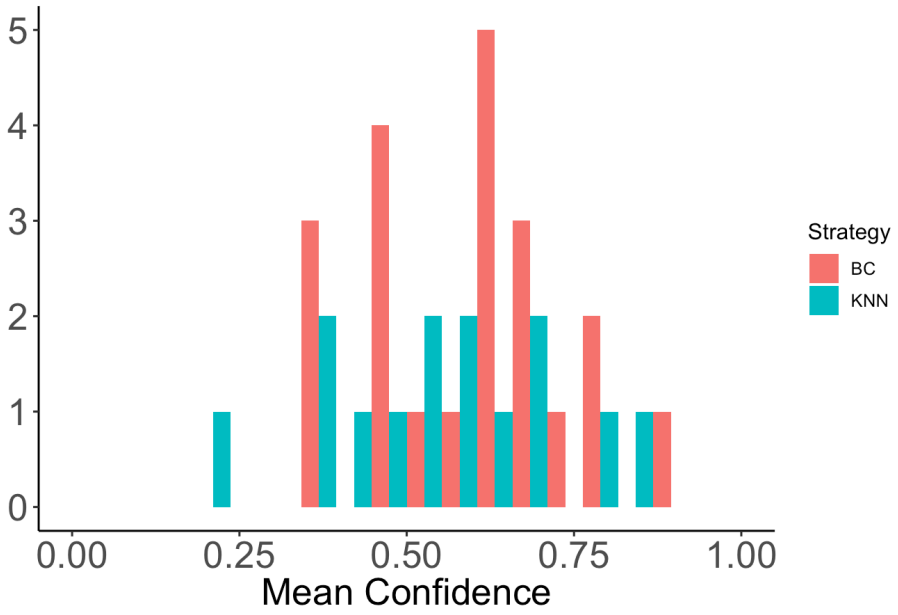

Figure S3. Average confidence (possible range = [0,1]) by subject and strategy. We did not find substantial evidence for a significant difference in the means or distributions across the two strategies (*p*s > 0.3).

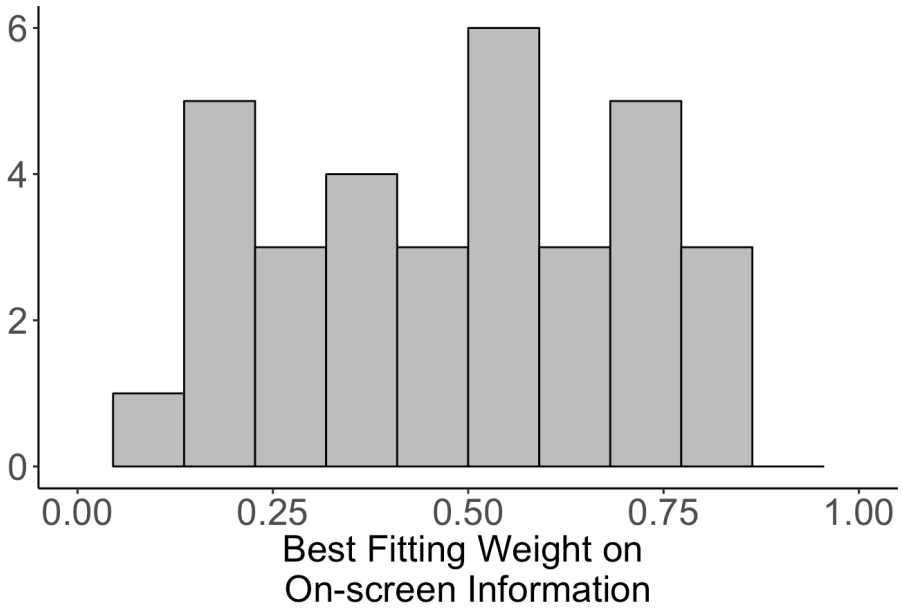

Fig S4. Distribution of best fitting weights on the on-screen information.

Alternate Beauty Contest Model

As mentioned in the main text, we compared subjects' guesses to a more complicated version of the BC model (CBC = Complex Beauty Contest). Specifically, we used the same leave-one-out cross-validation procedure specified in the main text. For each (left-out) trial, we estimated the following linear regression model:

$$CBC\text{: } Guess \sim \beta_0 + \beta_1 IMDb$$

Then, using the IMDb rating for the left-out trial, we estimated $\widehat{Guess}$ and recorded the error ($\widehat{Guess} - Guess$). The MSE across all trials was compared to the MSE for the KNN strategy for each subject. Unsurprisingly, more subjects were classified as CBC (24) than BC (21), which reduced the number of KNN subjects from 14 to 11. Table S1 shows the categorization of subjects using the new sorting procedure.

We also refit the second-block models using this categorization. The results are quite similar. Specifically, we see that subjects' actual weights on the onscreen information are strongly positively correlated with the optimal weights, $r(33) = 0.413$, $p = 0.014$. However, the distribution of best-fitting models differed quite a bit from the original analyses (as seen in Table S2). Specifically, 13 subjects were best fit with a model that includes primacy, while 16 were best fit with a model that includes similarity, 16 were best fit with a model that includes duration, and 21 were best fit with a model that includes recency.

Table S1. Revised strategy sorting and k-means classification

					Original		Revised	
Group	Current Film	Previous Films	Estimates	WTPs	KNN	BC	KNN	CBC
1	58%	**29%**	2%	11%	9	11	8	12
2	**80%**	12%	2%	6%	5	10	3	12

Table S2. Comparison of second-block models with original and revised sorting

Model	Number of subjects fit best by this model (original)		Number of subjects fit best by this model (revised)	
	KNN	BC	KNN	CBC
P	1	0	0	5
R	1	4	3	4
D	2	4	1	1
S	1	6	0	0
PR	0	2	1	0
PD	0	0	0	0
RD	0	1	2	1
PS	0	1	1	1
RS	2	0	0	2
DS	1	1	1	3
PRD	2	1	1	0
PRS	0	0	0	2
PDS	0	0	0	1
RDS	2	1	0	4
PRDS	2	0	1	0

List of Films

300
13 Going on 30
28 Days Later
50 First Dates
500 Days of Summer
8 Mile
A Bug's Life
A Christmas Carol
Ace Ventura: Pet Detective
Airplane!
Alice in Wonderland
Aliens vs. Predator 2
Alvin and the Chipmunks
Amelie
American Beauty
American Gangster
American Pie
American Psycho
Anchorman 2: The Legend Continues
Austin Powers: International Man of Mystery
Avatar
Bad Grandpa
Bad Santa
Batman Begins
Billy Madison
Black Hawk Down
Blade Runner
Borat
Brokeback Mountain
Burn After Reading
Cars
Casino Royale
Catch Me If You Can
Charlie and the Chocolate Factory
Chicken Run
Children of Men
Chocolat
Citizen Kane
Click
Cloverfield
Constantine
Couples Retreat
Crank
Crash

Deuce Bigalow: Male Gigolo
District 9
Django Unchained
Donnie Darko
Dumb and Dumber
Dumb and Dumberer: When Harry Met Lloyd
Elf
Eternal Sunshine of the Spotless Mind
Eurotrip
Exit Through the Gift Shop
Fantastic Four
Fargo
Fight Club
Final Fantasy: The Spirits Within
Finding Nemo
Friday Night Lights
Gangs of New York
Gladiator
Gone With The Wind
Gran Torino
Groundhog Day
Happy Gilmore
Harold & Kumar Go To White Castle
Harry Potter and the Chamber of Secrets
Harry Potter and the Goblet of Fire
Harry Potter and the Half-Blood Prince
Harry Potter and the Order of the Phoenix
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Sorcerer's Stone
He's Just Not That Into You
Hotel Rwanda
How the Grinch Stole Christmas
How to Lose a Guy in 10 Days
I Am Legend
I Heart Huckabees
Inglorious Basterds
Into the Wild
Iron Man
Jarhead
Jumper
Kill Bill: Vol. 1
Kingdom of Heaven
Knocked Up
Kung Fu Panda
Legally Blonde
Little Miss Sunshine

Mars Attacks!
Mean Girls
Meet the Fockers
Meet the Parents
Memento
Men in Black
Million Dollar Baby
Minority Report
Miss Congeniality
Miss Congeniality 2
Mission: Impossible
Monsters, Inc.
Munich
My Neighbor Totoro
Not Another Teen Movie
O Brother, Where Art Thou?
Ocean's Eleven
Pan's Labyrinth
Pearl Harbor
Peter Pan
Pineapple Express
Pirates of the Caribbean: At World's End
Pirates of the Caribbean: Dead Man's Chest
Pirates of the Caribbean: The Curse of the Black Pearl
Pulp Fiction
Quantum of Solace
Remember the Titans
Requiem for a Dream
Reservoir Dogs
Robots
Saving Private Ryan
Saw
Scream
Serenity
She's the Man
Sherlock Holmes
Shooter
Shrek
Signs
Sin City
Skyfall
Snatch.
Snowpiercer
Speed
Spider-Man
Spirited Away

Star Trek
Star Trek Into Darkness
Star Wars Episode I: The Phantom Menace
Star Wars: Episode II - Attack of the Clones
Step Brothers
Stranger Than Fiction
Super Size Me
Superbad
Taken
Tenacious D in The Pick of Destiny
The 40 Year-Old Virgin
The Aviator
The Big Lebowski
The Boondock Saints
The Bourne Identity
The Bourne Ultimatum
The Chronicles of Narnia: The Lion, the Witch and the Wardrobe
The Darjeeling Limited
The Dark Knight
The Departed
The Exorcist
The Fast and the Furious
The Girl with the Dragon Tattoo (Swedish version)
The Green Mile
The Hangover
The Hitchhiker's Guide to the Galaxy
The Incredible Hulk
The Incredibles
The Italian Job
The Lego Movie
The Life Aquatic with Steve Zissou
The Lion King
The Lord of the Rings: The Fellowship of the Ring
The Lord of the Rings: The Return of the King
The Lord of the Rings: The Two Towers
The Matrix
The Matrix Reloaded
The Passion of the Christ
The Polar Express
The Pursuit of Happyness
The Ring
The Santa Clause 2
The Shawshank Redemption
The Simpsons Movie
The Wolf of Wall Street
There Will Be Blood

Titanic
Toy Story
Trainspotting
Tropic Thunder
Tucker and Dale Vs. Evil
United 93
Up
V for Vendetta
Vertigo
Watchmen
Wayne's World
Wedding Crashers
Wild Hogs
X-Men
X-Men Origins: Wolverine
X-Men: The Last Stand
Zombieland
Zoolander