# Cognitive load selectively interferes with utilitarian moral judgment

Joshua D. Greene, Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, Jonathan D. Cohen

December 17, 2022

## Contents



## Introduction

In this project I will be attempt to reproduce the main finding of Greene et al 2008, who investigated the effect of cognitive load on moral decision making. Greene et al 2008, presented 40 vignettes describing moral dilemmas to 82 subjects mostly undergraduates at Princeton. Participants were asked tomake either deontological (rule based) or utilitarian (consequence based) choices for each dilemma. The study utilized a within-subjects design to manipulate cognitive load (i.e. all subjects made decisions in the presence and absence of load). Their primary hypothesis was that utilitarian choices would be specifically susceptible to cognitive load interference. To test their primary hypothesis the authors analyzed reaction times (RTs)for a subset (12) of these dilemmas and ran a mixed effects model to examine load and moral choice type on RT. they reported no main effect of load and a "marginally significant" (i.e. not signifcant) main effect of choice type. Critical to their hypothesis - they observed an interaction between load and choice type, such that participants took longer to make utilitarian responses, but only under conditions of cognitive load. This finding, in line with their hypothesis, they claim suggests support for the theory that utilitarian choices require more cognitive processing or use more cognitive effort.

I am attempting to replicate the primary finding of this paper, that the effect of load on RTs is selective for utilitarian vs. non-utilitarian decision making. This would be a mixed effects linear model looking at the effect of moral choice and load condition on response time, with subject included as a random effect.

### Justification for choice of study

I choose to replicate this study because it is a seminal piece in the moral psychology literature, that still shapes how investigators in this area conceptualize the cognitive mechanisms underlying moral decision-making. This work - which builds on previous work by Greene in 2001 - is situated within a dual-process decision making framework, and these results are construed as evidence supporting the existence of so-called "System 1" and "System 2" modes of cognition. Greene's work is highlighted strongly, for example, in Daniel Kanheman's "Thinking Fast Thinking Slow" (Kanheman, 2011). In particular, this work bridged the fields of judgement and decision-making and moral reasoning to create a lasting (and increasingly dominant) theory that moral choices arise from domain-general decision-making processes.

### Anticipated challenges

A primary challenge I anticipate is collecting response times that are comparable in range to those reported by Greene and colleagues. In the target paper, mean +/- 1SD was approximately 5.8 to 6s. These RTs seem implausible given that the vignettes are long text strings that require considerably longer than 10s to read when static and >40s when scrolling. The authors do not report important details about how RTs were collected, such as whether the window for measuring them began after subjects indicated they'd finished reading the vignette). As written, it would seem that RTs reflected the period between trial onset (i.e. the start of the streaming text vignette) and subject response. I will attempt to replicate the study as written, notwithstanding the vagueness regarding RT measurement in the original paper.

Finally, I will be conducting my replication online rather than in-person which can introduce other differences such as demographic differences. I do not anticipate differences a-priori due to this difference but it is entirely possible that in person experimental administration can have effects on moral decision-making broadly.

### Links

Project repository (on Github):

Original paper (as hosted in your repo):

https://github.com/lynde-m-folsom/Replications/blob/main/Greene_Rep/Original_paper/Greene_Cohen_Cog2007.pdf

## Methods

### How we intend to replicate

First, I will design a task that is close enough to the original to provide a meaningful test of replication. In the original study, The dilemmas were presented as scrolling text across a screen with the cognitive load manipulating comprising a number line also scrolling beneath the text, but at a different rate.This replication will be conducted online so the first challenge is replicating what the subjects experienced in the original study administration. To that effect, I will use a modified Qualtrics survey, utilizing custom CSS code to create the scrolling text effects reported in the paper. Considering the paper uses 40 vignettes but only analyzes 12 in the findings they report, we will use the 12 included in their analysis, breaking them evenly between load/no load conditions (order of load/no load blocks randomized by participant).

After piloting the design among peers and lab-mates. I will acquire the data in 82 subjects to follow the original study design. After gathering data, I will calculating the mean and standard deviations for RTs separately for each level of the load condition. Per the original paper, I will exclude responses that do not fall within 2 standard deviations of the (load condition level-specific) mean. Finally I will fit a series of linear mixed models to test the main hypothesis of Greene et al 2008. Below, I detail each of these models.

Model 1: RT ~ choice + (1|subject). This model tests for a main effect of choice type (deontological vs. utilitarian) on RT. For reference, the original study reported a "marginally significant" ($p = 0.053$) effect.

Model 2: RT ~ load + (1|subject).This model tests for a main effect of cognitive load on RT. For reference the original study did not find a main effect of load.

Model 3: RT ~ load + choice + load * choice + (1|subject). In this model, both first order terms (choice, load) and a second order interaction term (choice * load) were included as predictors of RT. This is the model employed for inference in the original paper.

### Differences from original study

One major difference is that I will be conducting this study online rather than in person. This means my population will not be primarily Princeton undergraduates, but instead may include a more diverse and demographically representative sample of subjects. Of note, it is possible replication sample may differ from the original sample in more "basic" aspects of cognition (e.g. reading speed). It is possible that such differences could "read out" as (artifactual) differences in moral decision-making under load.

Another major difference is that for the original study 40 dilemmas were presented and only 12 were included for the results. In this study I will only present the 12 that were used for their findings. This change was made to lower subject burden and make collection of an appropriate sample size tractable given constraints related to the nature of this replication (i.e. a class project).

Finally our study sample comprises 72 participants rather than our planned sample size of 82 due to a data acquisition error for 10 participants.

### Measure of success

The measure of success here will be weather we can replicate the main finding that cognitive load selectively interferes with utilitarian choices. This would be demonstrated by a significant result for the load * condition predictor in model three (described above).

## Results

### Data preparation

Data preparation following the analysis plan. #Library

### Setup

### Bring in Data & organize it

- Read in data

- Select the variables (IVs/DVs)
- Remove pilot data & Qualtrics headers
- Pivot long
- Trim RT to 2SD of mean
- Look at the histogram of RT
- Look at the violins of choice/load RT

```
#### Import data
raw_data <- read_survey("Greene_08_Rep_121422.csv", legacy = TRUE) %>%
  clean_names()
```

```
-- Column specification --------------------------------------------------------
cols(
  .default = col_character()
)
i Use `spec()` for the full column specifications.
```

```
#### Data exclusion / filtering
#grab the variables I want
df.data <- select(raw_data,
                  "prolific_id",
                  contains("l1"),
                  contains("l2"),
                  -contains("page_submit"),
                  -contains("last_click"),
                  -contains("click_count"))
#remove the pilot & qualtrics headers that I don't want
df.data <- df.data %>%
  filter(prolific_id !="beth")%>%
  filter(prolific_id !="kate") %>%
  filter(prolific_id !="ImportId") %>%
  filter(prolific_id !="test001_JWB") %>%
  filter(prolific_id !="asdf") %>%
  filter(prolific_id != '{"ImportId":"QID33_TEXT"}')
#### Prepare data for analysis - create columns etc.
df.datalong <- df.data %>%
  group_by("prolific_id") %>%
  pivot_longer(
    cols = !"prolific_id",
    names_to = c("trial", "load"),
    names_sep = "_",
    values_to = "choice")
```

```
Warning: Expected 2 pieces. Additional pieces discarded in 24 rows [1, 2, 3, 4,
5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
df.datalong <- df.datalong %>%
  filter(choice != "prolific_id" )
## Pesky row that Anna helped move
odd.ind <- seq_len(nrow(df.datalong)) %% 2
df.datalong.odds <- df.datalong[odd.ind == 1, ]
df.datalong.evens <- df.datalong[odd.ind == 0, ]

df.datalong.odds$bin.choice <- df.datalong.evens$choice

df.datalong
```

```
# A tibble: 1,728 x 4
   prolific_id              trial load  choice
   <chr>                    <chr> <chr> <chr>
 1 60bc16d1f9aef5318d50167d t1    l1    1
 2 60bc16d1f9aef5318d50167d t1    l1    40.767
 3 60bc16d1f9aef5318d50167d t2    l1    1
 4 60bc16d1f9aef5318d50167d t2    l1    49.667
 5 60bc16d1f9aef5318d50167d t3    l1    1
 6 60bc16d1f9aef5318d50167d t3    l1    53.364
 7 60bc16d1f9aef5318d50167d t4    l1    2
 8 60bc16d1f9aef5318d50167d t4    l1    56.266
 9 60bc16d1f9aef5318d50167d t5    l1    1
10 60bc16d1f9aef5318d50167d t5    l1    44.464
# ... with 1,718 more rows
```

```
df.datalong.odds
```

```
# A tibble: 864 x 5
   prolific_id              trial load  choice bin.choice
   <chr>                    <chr> <chr> <chr>  <chr>
 1 60bc16d1f9aef5318d50167d t1    l1    1      40.767
 2 60bc16d1f9aef5318d50167d t2    l1    1      49.667
 3 60bc16d1f9aef5318d50167d t3    l1    1      53.364
 4 60bc16d1f9aef5318d50167d t4    l1    2      56.266
 5 60bc16d1f9aef5318d50167d t5    l1    1      44.464
 6 60bc16d1f9aef5318d50167d t6    l1    1      48.513
 7 60bc16d1f9aef5318d50167d t1    l2    2      53.741
 8 60bc16d1f9aef5318d50167d t2    l2    2      41.006
 9 60bc16d1f9aef5318d50167d t3    l2    1      45.175
10 60bc16d1f9aef5318d50167d t4    l2    1      47.43
# ... with 854 more rows
```

```
new.df.datalong <- df.datalong.odds

df.datalong <-new.df.datalong %>%
  rename(rt = "bin.choice")

df.datalong <-df.datalong %>%
  mutate(rtnum = as.numeric(rt)) %>%
  mutate(choicenum = as.factor(choice))
```

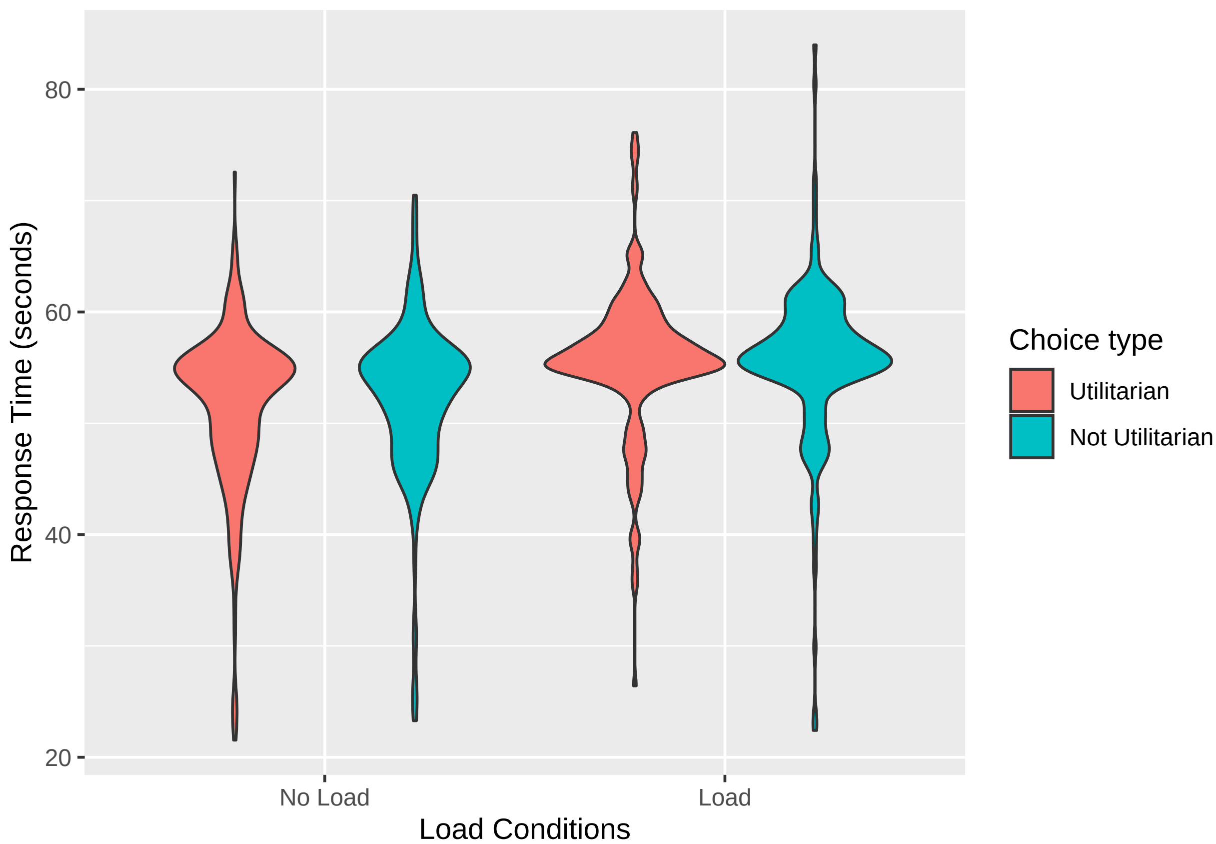
```r
## RT Trimming like in original paper

sumsdfdatalong <- df.datalong %>%
  group_by(load) %>%
  summarize(rtmean = mean(rtnum),
            stdrtmean = 2*sd(rtnum),
            minrt = rtmean-stdrtmean,
            maxrt = rtmean+stdrtmean)

minrt <- sumsdfdatalong$minrt
maxrt <- sumsdfdatalong$maxrt

df.datalongtrim <- df.datalong %>%
  group_by(load) %>%
  filter(rtnum > minrt) %>%
  filter(rtnum < maxrt)

## Lets see the data so far
df.datalongtrim %>%
  ggplot(mapping = aes(x = load,
         y = rtnum,
         fill = choicenum
        ))+
  geom_violin()+
  labs(x = "Load Conditions",
       y = "Response Time (seconds)")+
  scale_x_discrete(labels=c("l1" = "No Load",
                            "l2" = "Load"))+
  scale_fill_discrete(name="Choice type",
                 labels=c("1" = "Utilitarian",
                          "2" = "Not Utilitarian"))
```




**Key analysis**

- 3 linear models (participant as random effect)
  - M1: RT ~ Choice
  - M2: RT ~ Load
  - M3: RT ~ Load * Choice
- Anova model comparisons
- Plot the results

```
#modeling choice
m1<- lmer(rtnum ~choicenum +(1|prolific_id),
          data = df.datalongtrim)
reportm1<- report(m1)
summary(reportm1)
```

```
We fitted a linear mixed model to predict rtnum with choicenum. The model included prolific_id as rando

  - The effect of choicenum [2] is statistically non-significant and positive (beta = 0.76, 95% CI [-0.
```

```r
#modeling load
m2 <- lmer(rtnum ~load +(1|prolific_id),
           data = df.datalongtrim)
reportm2 <- report(m2)
summary(reportm2)
```

```
We fitted a linear mixed model to predict rtnum with load. The model included prolific_id as random eff

  - The effect of load [l2] is statistically significant and positive (beta = 3.59, 95% CI [2.89, 4.29]
```

```r
#model the load*choice and rt
m3 <- lmer(rtnum ~load *choicenum +(1|prolific_id),
           data = df.datalongtrim)
reportm3 <- report(m3)
summary(reportm3)
```

```
We fitted a linear mixed model to predict rtnum with load and choicenum. The model included prolific_id

  - The effect of load [l2] is statistically significant and positive (beta = 4.01, 95% CI [3.07, 4.95]
  - The effect of choicenum [2] is statistically non-significant and positive (beta = 0.82, 95% CI [-0.
  - The interaction effect of choicenum [2] on load [l2] is statistically non-significant and negative
```

```r
#looking at model comparisons
anova(m2, m3)
```

```
refitting model(s) with ML (instead of REML)

Data: df.datalongtrim
Models:
m2: rtnum ~ load + (1 | prolific_id)
m3: rtnum ~ load * choicenum + (1 | prolific_id)
   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
m2    4 4999.7 5018.4 -2495.8   4991.7
m3    6 5001.2 5029.2 -2494.6   4989.2 2.4809  2     0.2893
```

```r
anova(m1, m2)
```

```
refitting model(s) with ML (instead of REML)

Data: df.datalongtrim
Models:
m1: rtnum ~ choicenum + (1 | prolific_id)
m2: rtnum ~ load + (1 | prolific_id)
   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
m1    4 5092.3 5111.1 -2542.2   5084.3
m2    4 4999.7 5018.4 -2495.8   4991.7 92.693  0
```

```r
anova(m1, m3)
```

```
refitting model(s) with ML (instead of REML)
```

```
Data: df.datalongtrim
Models:
m1: rtnum ~ choicenum + (1 | prolific_id)
m3: rtnum ~ load * choicenum + (1 | prolific_id)
   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
m1    4 5092.3 5111.1 -2542.2   5084.3
m3    6 5001.2 5029.2 -2494.6   4989.2 95.174  2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#save model outputs to make a figure
means <- estimate_means(m3)
```

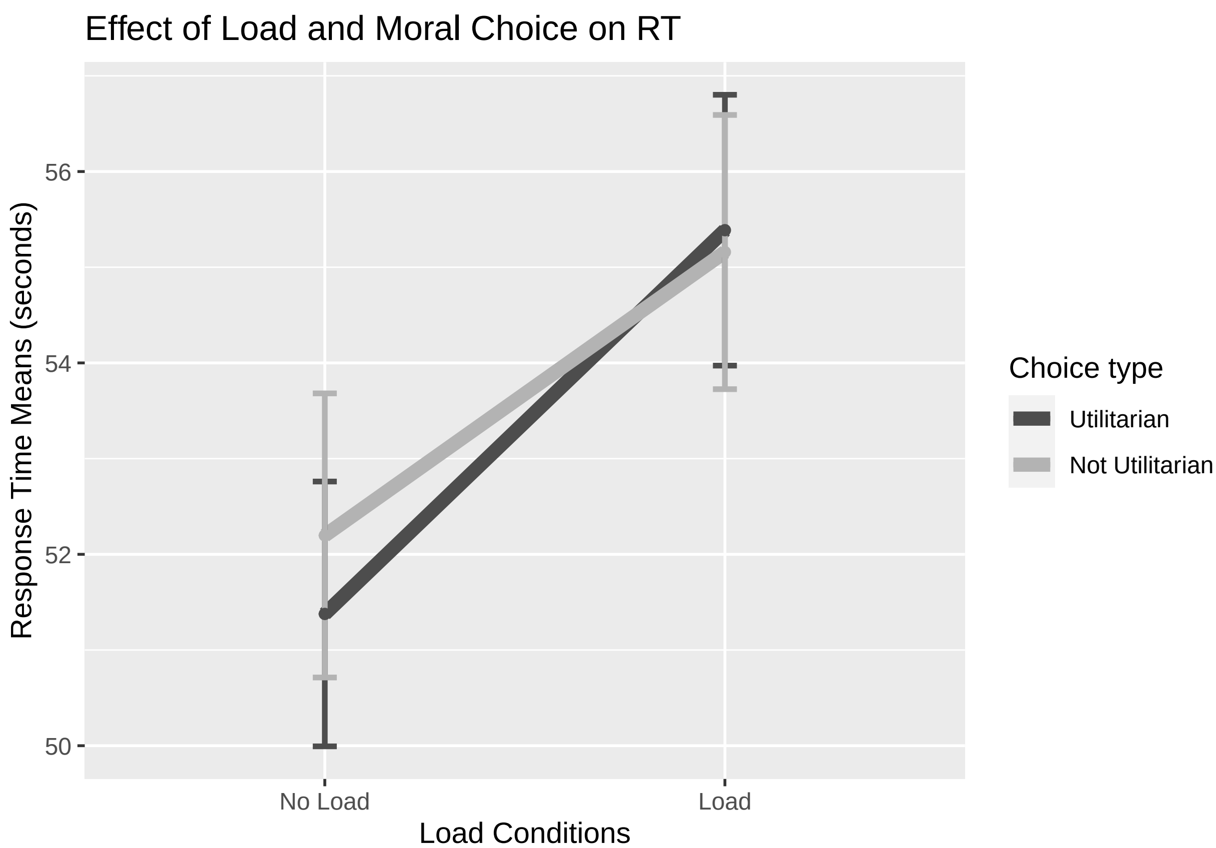
```
We selected `at = c("load", "choicenum")`.
```

```
as.factor(means$load) #this will make the graphing easier
```

```
[1] l1 l2 l1 l2
Levels: l1 l2
```

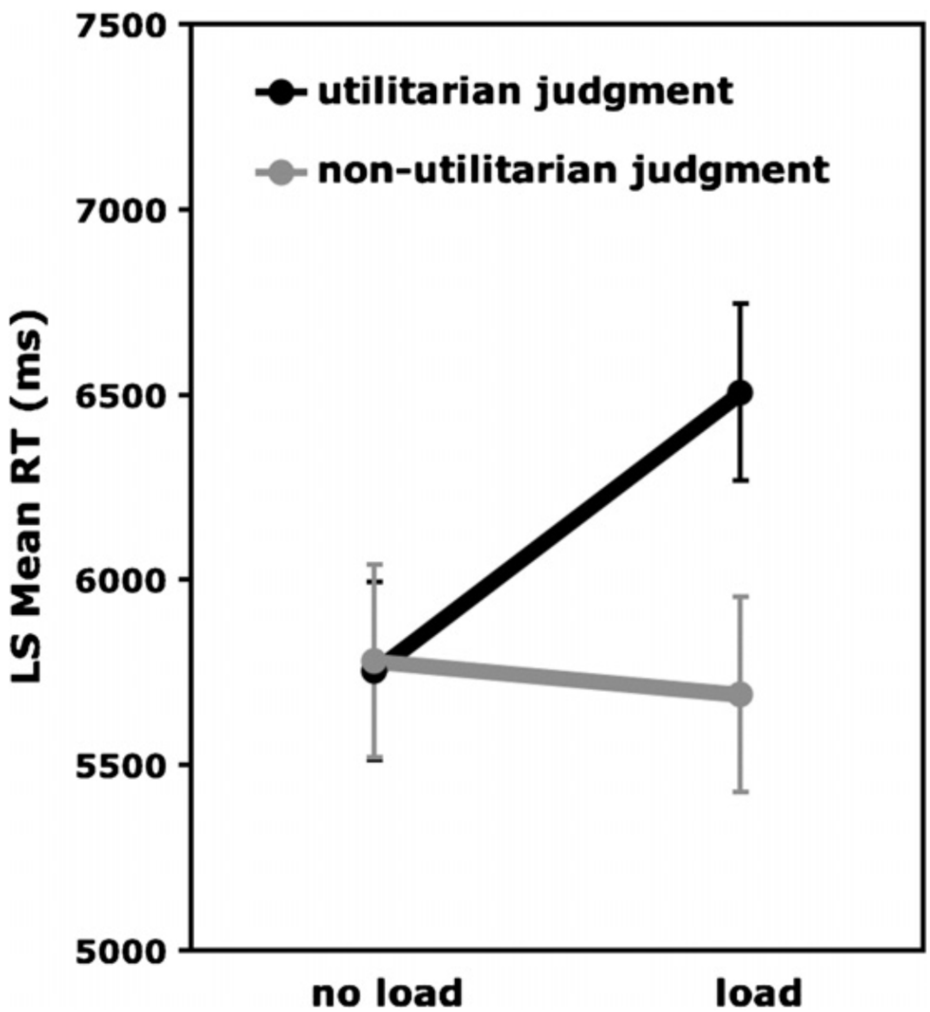
```
means %>%
  ggplot(mapping = aes(x = load,
                       y = Mean,
                       group = choicenum,
                       color = choicenum
                        ))+
  geom_line(linewidth = 2.5)+
  geom_errorbar(aes(ymin=CI_low, ymax=CI_high), width=.06, size = 1)+
  geom_point()+
  labs(title = "Effect of Load and Moral Choice on RT",
       x = "Load Conditions",
       y = "Response Time Means (seconds)")+
  scale_x_discrete(labels=c("l1" = "No Load",
                            "l2" = "Load"))+
  scale_colour_grey(name="Choice type",
                  labels=c("1" = "Utilitarian",
                           "2" = "Not Utilitarian"),
                  start = 0.3,
                  end = 0.7)
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



Effect of Load and Moral Choice on RT
Response Time Means (seconds)
56
54
52
50
No Load
Load
Load Conditions
Choice type
Utilitarian
Not Utilitarian

**Compare to the original figure**




```
#fig1<-include_graphics("~/F.Replications/Replications/Greene_Rep/Data/Figures/Greene_Fig1.png")
#fig1
```

## Discussion

### Summary of Reproduction Attempt

We first sought to determine whether the type of moral judgment (utilitarian/ not utilitarian) influenced RT during moral decision-making. In the original paper, the authors reported no significant effect for judgement type (F(1, 71.7) = 3.9, p = .052). Consistent with the original study, we did not find a significant effect (see "model 1"):

```
(beta = 0.76, 95% CI [-0.10, 1.61], t(791) = 1.73, p = 0.083, Std. beta = 0.11)
```

Original authors reported no main effect of load on RT ($F(1, 83.2) = 2.29$, $p = .13$; Greene, 2008). In contrast, however, we found that load significantly increased RTs (see “model 2”):

```
(beta = 3.59, 95% CI [2.89, 4.29], t(791) = 10.10, p < .001; Std. beta = 0.50, 95% CI [0.41, 0.60])
```

Further, I found a large difference in mean RT mean between the original study and this replication (original= 5.8s, current study = 52s).

Original work reports “predicted interaction between load and judgment ($F(1, 62.9) = 8.5$, $p = .005$).” By contrast, critically, I did not find a significant interaction between choice type and load on response times (see model 3) :

```
(beta = -1.05, 95% CI [-2.50, 0.40], t(789) = -1.42, p = 0.155, Std. beta = -0.15)
```

Due to these findings, in particular, the lack of an interaction between choice and load, we do not consider this a successful replication.

### Commentary

Considering the failure to replicate the effect, follow up questions I have for this study are as follows.

It is not clear to me that the “load” condition is introducing cognitive load rather than visual interference. In lit cognitive load literature often cites that tasks with conflicting sensory stimuli introduce sensory noise which I am not sure can be considered “cognitive load” (Tomlin, 2015). A better load condition may be something like an auditory odd-ball paradigm or n-back task in which a different cognitive modality is being used simultaneously rather than dual sensory processing(Hidaka, 2015).

Further, the scrolling text makes for a challenging task without introducing a load component, I would like to see how RT and judgment is effected by introducing scrolling text.

The content of the dilemmas are very similar. I believe this introduces a possible meta-cognitive factor that can reduce the cognitive load of the task by responding in a “heuristic” way (Aguinis, 2014).

Finally, the dilemmas all follow a similar model based off the trolley problem but with extreme consequences. I’m unsure the ecological validity of running graphic moral dilemmas to represent moral choices. These dilemmas are based off of philosophical thought experiments designed to solicit specific hypothetical responses not to solicit realistic responding. By adding graphic elements to the task I’m unsure if we get closer to understanding the relationship between cognitive processes and moral decision making.

### References


Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, *17*(4), 351–371. https://doi.org/10.1177/1094428114547952

Bago, B. (2022). Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. *Nature Human Behaviour*, *6*.

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801. https://doi.org/10.1037/xge0000533

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Hidaka, S., & Ide, M. (2015). Sound can suppress visual perception. *Scientific Reports*, *5*(1), 10483. https://doi.org/10.1038/srep10483

Tomlin, D., Rand, D. G., Ludvig, E. A., & Cohen, J. D. (2015). The evolution and devolution of cognitive control: The costs of deliberation in a competitive world. *Scientific Reports*, *5*(1), 11002. https://doi.org/10.1038/srep11002