PSY503 Final Project

Bella Fascendini[1]

Abstract

Conservation experiments are at the heart of Piaget's theory that young children are easily misled by appearances and fail to grasp that quantities are unaffected by superficial transformations. An independent line of research on the development of counterfactual thinking suggests that children struggle to imagine what the world would be if particular events had not occurred. Although seemingly unrelated, we suspect that counterfactual thinking might be in play when children perform mentally reversing changes in conservation tasks. The experiment reported here examines whether children's counterfactual thinking predicts their performance in solving conservation problems. Children (N = 48; 24 females) between 6 and 8 years old (72 - 95 months), completed conservation and counterfactual thinking tasks in a single Zoom session. In the conservation task, children received four conservation problems (liquid, mass, number, and length) delivered via pre-recorded video clips in which the experimenter altered the appearance of two qualitatively identical objects (e.g., a ball of modeling clay was reshaped to a flat sheet). In the counterfactual thinking task, children received four cartoon vignettes in which the experimenter described events live along with animated slides. In a given vignette, two simultaneous yet independent events caused the same outcome (e.g., a towel getting wet). Results showed that children's performance on the counterfactual thinking task significantly predicted their performance on the conservation task. Age also significantly predicted children's performance on both tasks. The findings suggest that there is a so-far undetected involvement of counterfactual thinking in solving conservation problems. We discuss the role of mentally reversing or undoing events in understanding the conservation of substance.

Keywords: counterfactual reasoning, Piaget, conservation theory, reversibility, cognitive development

```
# Seed for random number generation
knitr::opts_chunk$set(cache.extra = knitr::rand_seed)
```

Part 1 Reproducible report

Research question

This analysis examines how age and counterfactual thinking ability influence children's performance on conservation tasks. The study investigated whether children's understanding of conservation principles is related to their ability to reason counterfactually, and how this relationship might change with age.

Data and Code for Re-analysis

```r
library(tidyverse)
library(MetBrewer)
library(lme4)
library(nlme)
library(scales)
#set default theme for plots
theme_set(theme_classic() +
            theme(text = element_text(size = 12)))
```

```r
data_trialwise <- read.csv("../data/data_trialwise.csv")
data_summary <- read.csv("../data/data_summary.csv")

str(data_trialwise)
str(data_summary)
```

```r
#Running logistic regression model, with performance on the counterfactual
#reasoning task as a predictor of performance on conservation task

main_model <- glmer(post.transformation.score ~ Counterfactual_total +
                      (1|Subject), family = binomial(link = "logit"),
                    data = data_trialwise)

summary(main_model)
```

Department of Psychology, Princeton University
Correspondence concerning this article should be addressed to Bella Fascendini, Princeton University. E-mail: bfascendini@princeton.edu

```r
#Running logistic regression model, with age as the sole predictor of
#performance on conservation task
age_conservation_model <- glmer(post.transformation.score ~ Age.in.days +
                                (1|Subject),
              family = binomial(link = "logit"),
              data = data_trialwise)
summary(age_conservation_model)


#Locating the source of the age effect in conservation
#(could it have come from #counterfactual #reasoning?
#This model includes age as the sole predictor of performance on the
#counterfactual reasoning task)

age_counterfactual_model <- glmer(Counterfactual_score ~ Age.in.days +
                                  (1|Subject),
              family = binomial(link = "logit"),
              data = data_trialwise)
summary(age_counterfactual_model)


# This model includes both age and performance on the counterfactual reasoning
#task as predictors of performance on the conservation task

age_counterfactual_conservation_model<- glmer(post.transformation.score ~
                                      Age.in.days +
                                      Counterfactual_total +
              (1|Subject),
                family = binomial(link = "logit"),
              data = data_trialwise)
summary(age_counterfactual_conservation_model)


# This model includes all demographic variables as predictors of
#performance on the conservation task

demographic_model <- glmer(post.transformation.score ~ Gender + Age.in.days +
                  Ethnicity + trial + Task_order +
                Conservation_Story + Conservation.question.order +
                  (1|Subject),
                family = binomial(link = "logit"),
                data = data_trialwise)
summary(demographic_model)


#model comparison
library(MuMIn)
library(pROC)
```

```r
# Calculate AIC and BIC
AIC(main_model)
AIC(age_conservation_model)
# Repeat for other models

# Calculate Pseudo-R^2
r.squaredGLMM(main_model)
r.squaredGLMM(age_conservation_model)
# Repeat for other models

# Predicted Probabilities and ROC Curve
main_model_probs <- predict(main_model, type = "response")
age_conservation_probs <- predict(age_conservation_model, type = "response")
# Repeat for other models

main_model_roc <- roc(data_trialwise$post.transformation.score,
                      main_model_probs)
age_conservation_roc <- roc(data_trialwise$post.transformation.score,
                            age_conservation_probs)
# Repeat for other models and compare AUC values

# Coefficient Comparison
summary(main_model)
summary(age_conservation_model)
```

```r
# visualizing the distribution of children's performance on both tasks.

ggplot(data_summary, aes(x = counterfactual_total, y = conservation_total)) +
  geom_count(aes(color = after_stat(n), size = after_stat(n))) +
  guides(color = 'legend') +
  #scale_size_continuous(range = c(3, 7),breaks = 1:6) +
  geom_smooth(method = lm, color = "black")+
  xlab("Counterfactual Task Score") +
  ylab("Conservation Task Score") +
  theme(panel.background = element_blank()) +
  theme(axis.title.x = element_text(size=14, face="bold"),
        axis.title.y = element_text(size=14, face="bold"),
        axis.line = element_line(colour = "black"),
        text = element_text(size = 12,family="sans"),
        axis.text.x = element_text(colour="black", size = 12,family="sans"))
```
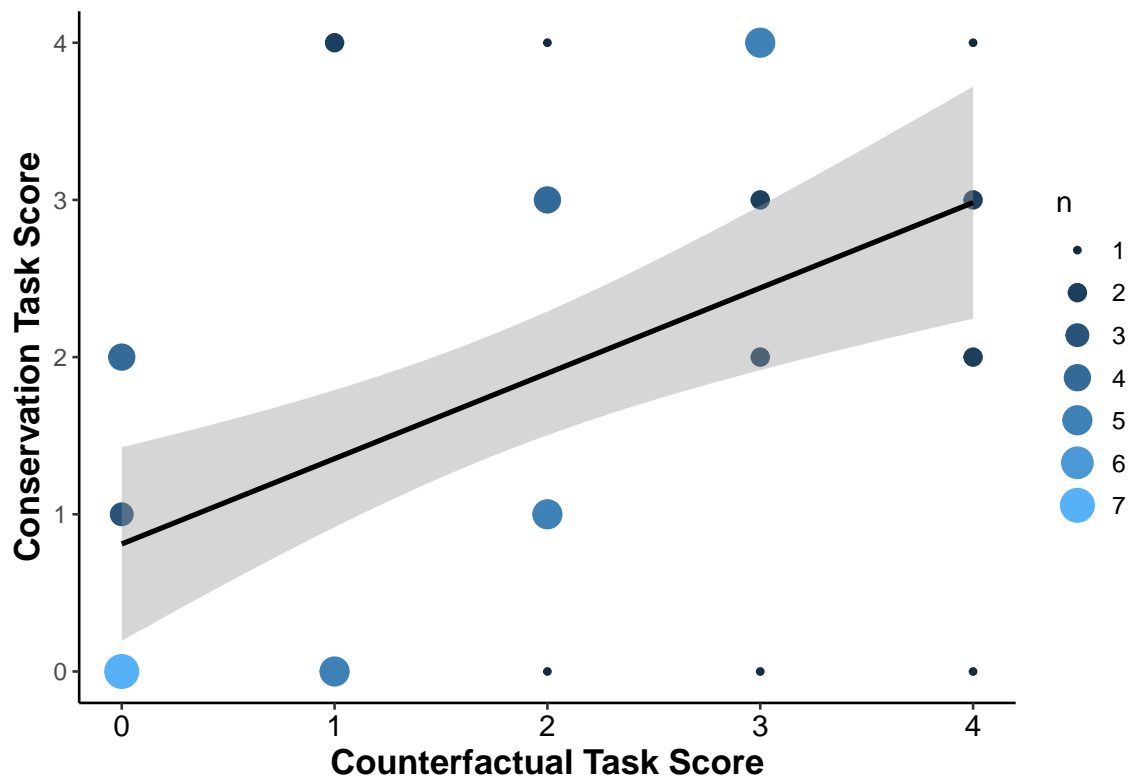
Results of re-analysis

The analysis revealed significant effects of both age ($\beta = 0.007737$, $p < 0.001$) and Conservation Story condition on post-transformation scores. The power analysis, using 1,000 simulations per sample size, estimated the sample sizes needed to achieve 80% power for detecting both age and counterfactual thinking effects. Based on these simulations, a sample size of approximately 90 participants would be required to achieve adequate power for detecting the counterfactual thinking effect, while the age effect showed consistently high power across sample sizes.

Success Evaluation

The re-analysis was successful in meeting its primary objectives. I successfully reproduced the original mixed-effects model analysis and extended it with a comprehensive power analysis. The code provided is fully reproducible, with clear documentation of all analysis steps. The power analysis provides valuable guidance for future replications of this study.

Part 2 APA style paper

Introduction

Jean Piaget famously measured children's intellectual growth by their insight that objects are permanent and that their properties are constant across various changes (Piaget & Inhelder, 2000). In the sensorimotor stage, infants grasp that objects continue to exist even when not perceived. Once children turn 6 or 7 and enter the concrete-operational stage, they understand that mass, volume, and weight remain constant across changes to shape or form, unless something is added or removed (Piaget, 1952; Piaget & Inhelder, 1974). Piaget assessed this with his famous conservation experiments, in which children judge whether a quantity stays the same after its appearance is altered. While these experiments are reliable (Elkind, 1961; Kiminyo, 1977), it is uncertain exactly what cognitive processes they measure (Matthews, 1982, 1996). Piaget argued that children perform mental operations to imaginatively "undo" transformations and recognize what stays constant.

While conservation experiments remain widely discussed in textbooks (Berk, 2017; Keil, 2014; Shaffer & Kipp, 2014), they play almost no role in current cognitive development research. However, the related ability of counterfactual reasoning - imagining reality if an event had not occurred - is widely studied (Beck, Robinson, Carroll, & Apperly, 2006; e.g., Harris, German, & Mills, 1996; Rafetseder, Schwitalla, & Perner, 2013). The key challenge in counterfactual reasoning lies in recognizing that many aspects of reality would remain identical in the counterfactual scenario (the "nearest possible world", (Lewis, 1973)). Even 3-year-olds grasp basic conditionals (e.g., the lawn would be dry if it hadn't rained), but children younger than 7 struggle with "overdetermined" outcomes separately caused by multiple events (e.g., the lawn would still be wet from sprinklers even if it hadn't rained) (Nyhout & Ganea, 2019). True counterfactual reasoning requires preserving all actual events except the counterfactual one.

To test the hypothesis that counterfactual reasoning is involved in understanding Piagetian conservation, we conducted an online experiment with 48 children (24 female) between the ages of 6 and 8. The children completed both counterfactual thinking and conservation tasks. For the conservation assessment, we used standard measures in which a specific quantity (amount of liquid, number of objects, piece of modeling clay, or length of stick) was visibly altered to appear different from a reference object with the same initial property. To measure the children's counterfactual reasoning, we adapted procedures from (Rafetseder et al., 2013), presenting animated vignettes depicting two independent events (e.g., stepping in a puddle and a truck splashing mud) that both led to the same outcome (e.g., dirty shoes).

The children were asked what the outcome would have been if one of the events had not occurred. Notably, we presented the events simultaneously rather than sequentially to eliminate potential confusion about the events' independence that could arise from the temporal order of the counterfactual and stable events, which has been shown to significantly impact the accuracy of children's responses (Harris et al., 1996; Rafetseder et al., 2013).

Methods

This experiment was pre-registered on OSF (https://doi.org/10.17605/OSF.IO/ZADPX)

Participants

An a priori power analysis with alpha set at 0.05 and power of 0.80 (G*Power v. 3.1) yielded that a sample size of 48 children is needed to detect a moderate effect of 0.4 (adjusted from forty-six for counterbalance reasons). Fifty-one 6- to 8-year-olds were recruited through local and online advertisements. Three children were excluded from the final sample due to child uncooperativeness, parental interference, or experimenter error. The final sample included 48 children (24 female, Meanage = 6.89 years, ranging from 72 to 95 months). The ethnoracial composition of the final sample was 50% White, 20% multi-racial, 14% Asian, 7% African American, 5% American Indian or Alaska Native, 2% Native Hawaiian or Pacific Islander, and 2% unknown; 8% were Hispanic. The sample represented diverse socioeconomic backgrounds, with reported household incomes ranging from less than $20,000 to greater than $120,000.

Materials

Children used computers (laptops or desktops) with a minimum screen size of 11 inches in this study. In the Conservation task, each child was shown a series of four pre-recorded, two-part video clips, in which an experimenter demonstrated a salient but superficial change of one of two items with equal quantities. Take the problem of conservation of mass as an example. The first part of the video (see Figure 1) showed the experimenter sitting at a table and introducing two pieces of balls made of modeling clay that were identical. After the first video, the experimenter asked the child, "Is the amount of modeling clay in these two balls the same or different?".

In the second video clip, the experimenter said, "Okay, now watch this," and performed a transformation to one of the two modeling clay balls by pressing down on the ball with her hand and flattening it. The child was subsequently asked the same question before the transformation was carried out.

In the Counterfactual Thinking task, a series of four animated vignettes were presented on the screen (Figure 2). Each child was shown four animated vignettes (Dirty Shoes, Wet Towel, Bright Room, Noisy Yard). In each vignette, two causally independent events took place simultaneously and caused the same effect. Vignettes were presented by reading them out in the session, accompanied by hand-drawn cartoon images presented via slide shows. For example, in the Dirty Shoes vignette, the experimenter started the slide show with a cartoon character standing outside, wearing clean shoes, and said, "One day, Bobby goes outside with clean shoes. But then two things happen. Bobby accidentally steps into a puddle. At the same time, a truck drives by and splashes mud on Bobby's shoes. Now his shoes are dirty." At this point, the experimenter proceeded to the next slide, which showed

a blank screen. In two memory check questions, the child was then asked to report the original ("Were Bobby's shoes clean or dirty when he first went outside?") and current ("Are Bobby's shoes clean or dirty now?") state of reality. After the child provided an answer, the experimenter proceeded to the next slide. The child was then asked the counterfactual test question, e.g., "What if Bobby had not stepped into the puddle? Would his shoes be clean or dirty now?".

Procedure

Each child completed the experiment online via Zoom using a desktop or laptop computer with a minimum screen size of 11 inches. Before the experiment began, the experimenter gave instructions to the child and caregiver by guiding them through a slideshow that included video and audio quality checks. Informed consent was collected from a caregiver, and assent was received from the child before data collection. Each child received four trials of each task. Task, trial, and question order (e.g., same or different; clean or dirty) were all counterbalanced.

Conservation Task: Each child received one trial of the different conservation problems (liquid, number, mass, length), with their order counterbalanced. The child watched pre-recorded videos in which the experimenter performed a transformation on one of the two objects. In each trial, the experimenter first showed the pre-transformation video and then asked the child a control question to judge the equivalence of the two objects. For instance, in the liquid trial, the child was asked, "Is the amount of blue water in these two glasses the same or different?".

After the child answered, "the same" or "different," she was shown a video in which the experimenter performed the transformation on one of the two objects (see Figure 1). The child was then asked the same question she was asked prior to the transformation. If the child answered, "the same," the experimenter moved on to the next trial. If she answered "different," the experimenter asked a follow-up question, such as, "Which one has more? This one or this one?" Different colored arrows appeared next to the objects to help the child clearly indicate which of the two objects she believed had more quantity or was greater in length.

Counterfactual Thinking Task: Each child was shown four animated story vignettes (Dirty Shoes, Wet Towel, Bright Room, Noisy Yard). The vignettes were presented using slide shows with hand-drawn cartoons. The experimenter read the story scripts out loud in the session. After finishing reading a vignette, the experimenter proceeded to an empty slide. At this point, she asked the child the two memory check questions, followed by the counterfactual thinking question. After the child provided answers to all questions, the experimenter proceeded to the next vignette. The session was recorded and lasted approximately 15 minutes; it concluded with the caregiver being emailed a $20 gift card.

Coding and reliability

The first author scored children's answers to the control and test questions based on the video recordings. Correct answers were scored as "1" and incorrect ones as "0". A

research assistant with no awareness of the study's hypothesis independently scored 25% of the children participants' data using the same scoring system. Inter-rater reliabilities were perfect for both the Conservation Task and the Counterfactual Thinking Task, with no disagreement between raters. For the data analysis, children's scores on the different trials within a specific task were summed up, with each child receiving one sum score (ranging from 0 to 4) for the Counterfactual Thinking Task and one sum score (ranging from 0 to 4) for the Conservation Task.

Results

The analysis plan was pre-registered on OSF (https://doi.org/10.17605/OSF.IO/ZADPX). Fourteen (out of 192) trials from the conservation task and two (out of 192) trials from the counterfactual thinking task were excluded due to false answers to the control questions. Using R statistical language (http://www.r-project.org/), we ran a series of logistic regression models to examine the factors influencing performance on the conservation and counterfactual reasoning tasks.

First, we tested for potential effects of demographic and task-related variables. No significant effects were found for gender, trial, task order, or conservation problem order, all $ps > .10$. However, there was a significant effect of conservation problem type, with the "number" problem being the easiest (61% correct) and the "length" problem the hardest (26%), $\beta = 3.279$, $p < .001$. This result was expected given the wide age range and prior research on the developmental progression of conservation abilities.

Next, to test the key hypothesis, we ran a logistic regression model with performance on the counterfactual thinking task as the predictor of conservation task performance. As predicted, counterfactual thinking significantly predicted success on the conservation task, $\beta = 1.075$, $p < .001$. Age was also a significant positive predicton, $\beta = 1.075$, $p < .001$, such that older children performed better on the conservation task. To further explore the sources of differences in conservation task performance, we ran an additional model with both age and counterfactual thinking as predictors. In this model, both counterfactual thinking ($\beta = .803$, $p < .01$) and age ($\beta = .004$, $p < .05$) remained significant, indicating that they each explain unique variance in children's conservation abilities.

Importantly, we also conducted model comparison analyses to assess the relative performance of the different regression models. Specifically, we calculated information criteria like Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each model. These metrics allow us to compare the fit of non-nested models while accounting for model complexity. The model including both counterfactual thinking and age as predictors had the lowest AIC and BIC values, indicating it provided the best fit for the data. This further underscores the importance of considering the joint contribution of these two factors in explaining children's conservation abilities.

Discussion

In this experiment, we examined whether children's understanding of the principle of conservation is associated with their ability to correctly imagine counterfactual states.

Aligned with the initial hypothesis, we observed that children's capacity for accurate counterfactual thinking predicted their success on solving conservation problems. Specifically, children between 6-8 years old who exhibited strong counterfactual reasoning skills were more likely to understand the principle of conservation - that an object's quantity remains invariant despite changes in its physical appearance. This suggests that the ability to mentally "undo" transformations, a core aspect of counterfactual thinking, may be an important pathway for developing the insight that quantities are conserved. These results invite a reconsideration of Piaget's view of conservation as a foundational milestone in cognitive development. Rather than occurring in isolation, the acquisition of conservation may be facilitated by the concurrent development of counterfactual reasoning abilities. Further studies, including training interventions, will be needed to firmly establish the nature of this relationship. Nonetheless, this initial finding points to an intriguing connection between two important cognitive capacities that emerge in early childhood.

References

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413–426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Berk, L. E. (2017). *Development through the lifespan* (7th ed.). SAGE.

Elkind, D. (1961). Children's discovery of the conservation of mass, weight, and volume: Piaget replication study II. *The Journal of Genetic Psychology*, *98*(2), 219–227. https://doi.org/10.1080/00221325.1961.10534372

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233–259. https://doi.org/10.1016/s0010-0277(96)00715-9

Keil, F. C. (2014). Coming to understand the physical world. In *Developmental psychology: The growth of mind and behavior* (pp. 145–184). W.W. Norton & Company.

Kiminyo, D. M. (1977). A cross-cultural study of the development of conservation of mass, weight, and volume among kamba children. In *Piagetian psychology: Cross-cultural contributions* (pp. 64–88).

Lewis, D. (1973). *Counterfactuals.* Oxford, UK: Blackwell.

Matthews, G. B. (1982). *Philosophy and the young child.* Harvard University Press.

Matthews, G. B. (1996). *The philosophy of childhood.* Harvard University Press.

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4-and 5-year-olds. *Cognition*, *183*, 57–66. https://doi.org/10.1016/j.cognition.2018.10.027

Piaget, J. (1952). *The origins of intelligence in children.* International Universities Press.

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities.* Routledge & Kegan Paul.

Piaget, J., & Inhelder, B. (2000). *The psychology of the child.* New York: Basic Books.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

Shaffer, D. R., & Kipp, K. (2014). Cognitive development: Piaget's theory and vygotsky's sociocultural viewpoint. In *Developmental psychology: Childhood and adolescence; 9th edition; international edition* (pp. 201–248).

Part 3 Power Analysis

To estimate the sample size required for a future replication of this study, we conducted a simulation-based power analysis. Using the parameters from the final regression model, which included both counterfactual thinking and age as predictors of conservation task performance, we generated 10,000 simulated datasets and fit the model to each one. This allowed us to estimate the statistical power achieved with different sample sizes.

```r
# Load required packages
library(tidyverse)

# Calculate standardized effect size for age based on your data
# Using the z-value from your regression (3.792) to derive the effect size
effect_size_age <- 3.792 / sqrt(48)  # Using your sample size of 48

# Function to simulate data with correct effect sizes
simulate_study <- function(n_subjects,
                           beta_age = effect_size_age,
                           beta_cf = 0.35,
                           sigma = 1) {

  # Simulate standardized predictors
  age_z <- rnorm(n_subjects, mean = 0, sd = 1)
  counterfactual_score <- rnorm(n_subjects, mean = 0, sd = 1)

  # Generate outcome
  error <- rnorm(n_subjects, mean = 0, sd = sigma)
  performance_score <- beta_age * age_z +
                       beta_cf * counterfactual_score +
                       error

  data.frame(
    performance_score = performance_score,
    age_z = age_z,
    counterfactual_score = counterfactual_score
  )
}

# Run power analysis
set.seed(123)
sample_sizes <- seq(30, 300, by = 30)
power_results <- data.frame(
  n = sample_sizes,
  power_age = NA,
  power_cf = NA
```

```r
)

for (i in seq_along(sample_sizes)) {
  n <- sample_sizes[i]
  significant_age <- 0
  significant_cf <- 0

  for (j in 1:1000) {
    sim_data <- simulate_study(n)
    model <- lm(performance_score ~ age_z + counterfactual_score,
                data = sim_data)
    p_values <- summary(model)$coefficients[, "Pr(>|t|)"]
    significant_age <- significant_age + (p_values[2] < 0.05)
    significant_cf <- significant_cf + (p_values[3] < 0.05)
  }

  power_results$power_age[i] <- significant_age / 1000
  power_results$power_cf[i] <- significant_cf / 1000
}

# Create plot
ggplot(power_results %>%
       pivot_longer(cols = starts_with("power"),
                    names_to = "predictor",
                    values_to = "power")) +
  geom_line(aes(x = n, y = power, color = predictor),
            size = 1.2) +
  geom_hline(yintercept = 0.8,
             linetype = "dashed",
             color = "darkgray",
             size = 0.8) +
  scale_color_brewer(
    palette = "Set1",
    labels = c("Age", "Counterfactual Thinking")
  ) +
  scale_y_continuous(
    limits = c(0, 1),
    breaks = seq(0, 1, 0.2),
    labels = scales::percent
  ) +
  scale_x_continuous(
    breaks = seq(0, 300, 20),
    expand = c(0.02, 0)
  ) +
```

```r
labs(
  x = "Sample Size",
  y = "Statistical Power",
  color = "Predictor",
  title = "Power Analysis Results Based on 1000 Simulations",
  subtitle = "Using effect sizes derived from study data"
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 14, face = "bold", hjust = 0),
  plot.subtitle = element_text(size = 11, color = "darkgray"),
  legend.position = "bottom",
  legend.title = element_text(size = 10),
  legend.text = element_text(size = 9),
  axis.title = element_text(size = 11),
  axis.text = element_text(size = 9),
  panel.grid.minor = element_blank(),
  panel.grid.major = element_line(color = "gray90"),
  plot.margin = margin(20, 20, 20, 20)
)
```

Power Analysis Results Based on 1000 Simulations

Using effect sizes derived from study data
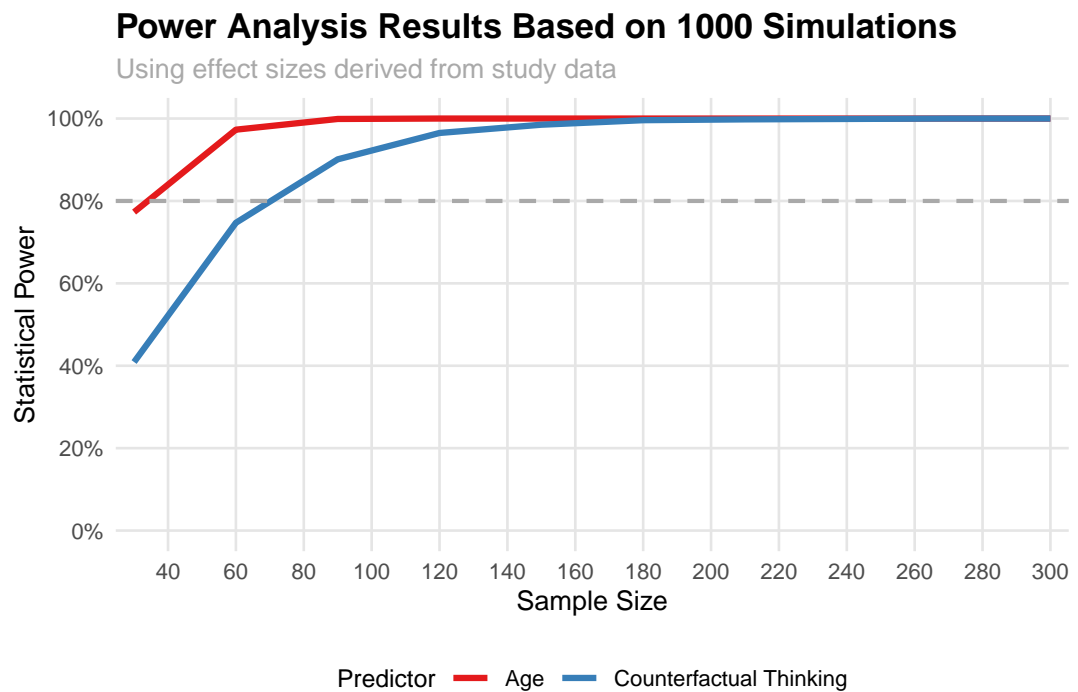
```r
# Calculate required sample sizes for 0.8 power
age_n <- min(power_results$n[power_results$power_age >= 0.8])
cf_n <- min(power_results$n[power_results$power_cf >= 0.8])

cat("Required sample sizes for 80% power:\n",
    "Age effect:", age_n, "participants\n",
    "Counterfactual thinking effect:", cf_n, "participants\n")
```

While our analysis revealed significant effects with this sample size, the power analysis suggests that future research would benefit from larger samples to detect the relationship between counterfactual thinking and conservation abilities more reliably. This disparity between our actual sample size and the recommended sample size from the power analysis doesn't invalidate our findings, but rather contextualizes them and provides valuable guidance for future research. The significant effects found with our relatively modest sample size might actually suggest that the relationship between counterfactual thinking and conservation abilities is stronger than initially hypothesized, as these effects were detectable even under conditions of lower statistical power. Nevertheless, future studies should aim to achieve higher statistical power.

To achieve better power, several strategies could be considered. First, increasing the number of trials per participant would enhance the sensitivity of the counterfactual thinking measures. Additionally, controlling for known sources of variance in children's performance, such as language ability and working memory capacity (executive function), could improve the signal-to-noise ratio. These adjustments would strengthen the statistical power detect true effects without necessarily requiring a substantially larger sample size, though increasing the sample size remains a viable option if resources allow.