

Behavioral Experiment Simulation Tool

Proprietary Software by Dr. Eugen Dimant

Generate publication-ready synthetic data for behavioral science experiments

Overview

The Behavioral Experiment Simulation Tool is a sophisticated platform that generates high-fidelity synthetic datasets for behavioral science research. Whether you're piloting a new experiment, preparing analysis code, or teaching research methods, this tool produces data that mirrors what you would collect from actual human participants.

Unlike simple random number generators, this system applies decades of survey methodology research to create responses that exhibit realistic psychological properties—including response styles, attention patterns, and treatment effects calibrated to your specifications.

Who Benefits from This Tool?

Academic Researchers

- **Pre-registration preparation:** Generate synthetic data to develop and test your analysis scripts before collecting real data
- **Power analysis:** Validate sample size calculations with data that matches your expected effect sizes
- **Grant applications:** Include preliminary analyses in proposals using realistic simulated data
- **Pilot testing:** Test your experimental design logic and identify issues before investing in data collection

Graduate Students and Postdocs

- **Methods training:** Learn data analysis techniques with realistic datasets that have known properties
- **Thesis preparation:** Develop analysis pipelines while awaiting IRB approval
- **Replication planning:** Generate data matching published effect sizes to plan replication studies

Course Instructors

- **Teaching materials:** Create datasets with specific statistical properties for classroom exercises
- **Exam development:** Generate realistic data for assessment questions
- **Student projects:** Provide synthetic data for students who cannot access real participant data

Industry Researchers

- **A/B test planning:** Simulate experiment outcomes before deploying to production
 - **UX research:** Generate survey response data for prototyping analysis dashboards
 - **Stakeholder presentations:** Demonstrate analysis approaches with representative data
-

How It Works

The tool offers **two input pathways** to accommodate different stages of the research process:

Pathway A: Upload Your Qualtrics Survey (.qsf)

If you already have a Qualtrics survey built, simply upload the .qsf file. The system automatically extracts:

- Experimental conditions from your BlockRandomizer or branch logic
- Dependent variables (Likert scales, sliders, matrices)
- Open-ended questions and their visibility rules
- Survey flow logic determining which questions each participant sees

Pathway B: Describe Your Experiment (Conversational Builder)

If you haven't built your survey yet — or prefer a faster setup — you can **describe your experiment in plain language**. The conversational builder guides you through:

1. **Study title and description:** What is your study about?
2. **Conditions:** Describe your experimental design in natural language. The system automatically detects:
 - Simple designs: "Treatment vs Control"
 - Factorial designs: "3 (Annotation: AI vs Human vs None) × 2 (Product: Hedonic vs Utilitarian)"
 - Numbered lists: "1. Low dose, 2. Medium dose, 3. High dose"
 - Any NxM crossed design with automatic condition generation
3. **Scales and DVs:** Describe your measures in paragraph or list format. The parser recognizes:
 - Standard scale specifications: "Trust scale (4 items, 1-7)"
 - Detailed academic format: "Perceived Quality (PQ): 3 items (7-point Likert; 1=low, 7=high)"
 - Known validated instruments: "BFI-10", "PANAS", "GAD-7", "PHQ-9"
 - Numeric measures: "Willingness to Pay (WTP): 1 item (open-ended numeric)"
 - Binary measures: "Manipulation check (Yes/No)"
4. **Open-ended questions:** Simply list your qualitative questions
5. **Research domain:** Select from 225+ research domains for persona-appropriate responses
6. **Sample size and effect sizes:** Configure your simulation parameters

The builder outputs the same structured design specification used by the QSF pathway, ensuring identical simulation quality regardless of input method.

Design Review

Regardless of input method, you review and adjust the detected design:

- Verify conditions, scales, and open-ended questions
- Edit names, scale ranges, and item counts inline
- Add or remove measures as needed
- Customize persona weights for domain-specific response patterns
- Set expected effect sizes (Cohen's d) with visual condition selectors

Generate and Download

The system produces a publication-ready CSV file containing:

- Participant IDs and condition assignments
 - Likert scale responses with realistic distributions
 - Unique open-ended text responses
 - Demographics and metadata
 - Quality metrics and validation flags
 - A comprehensive instructor report with statistical analyses, persona breakdowns, and effect size verification
-

The Science Behind the Simulation

Persona-Based Response Generation

Rather than generating random responses, the system assigns each simulated participant a "persona" based on survey methodology research. **These personas have been trained on hundreds of scientific insights from decades of research across the social and behavioral sciences**, including:

- **Behavioral Economics:** Trust, cooperation, altruism, fairness, reciprocity, risk preferences, loss aversion, framing effects
- **Social Psychology:** Social identity, group dynamics, conformity, prosocial behavior, intergroup relations, attitudes, persuasion
- **Cognitive Psychology:** Decision-making heuristics, cognitive biases, memory, attention, reasoning processes
- **Organizational Behavior:** Leadership, motivation, job satisfaction, team dynamics, workplace attitudes
- **Political Psychology:** Polarization, partisanship, civic engagement, media effects, political trust
- **Consumer Behavior:** Brand perception, purchase intent, product evaluation, advertising effectiveness
- **Moral Psychology:** Ethical judgment, values, moral emotions, fairness perceptions
- **Health Psychology:** Medical decisions, wellbeing, health behaviors, treatment preferences

This rich scientific foundation enables each persona to generate responses that align with documented human response patterns across these diverse research domains.

The personas reflect actual patterns observed in human respondents:

Engaged Responders (35%): High attention, thoughtful responses, full scale use. Based on Krosnick's (1991) "optimizers" who invest cognitive effort in providing accurate answers. These participants draw on genuine reflection about the topic at hand.

Satisficers (22%): Lower effort responses, tendency toward agreement, restricted scale range. Krosnick's research documented this common response strategy where participants provide acceptable rather than optimal answers.

Extreme Responders (10%): Consistent use of scale endpoints. Greenleaf's (1992) work identified this stable response style that varies across individuals.

Acquiescent Responders (8%): Strong agreement bias regardless of item content. Billiet & McClendon's (2000) studies documented this tendency to agree with statements.

Careless Responders (5%): Low attention, random patterns. Meade & Craig's (2012) research characterized these inattentive participants common in online samples.

Effect Size Calibration

Treatment effects are calibrated using Cohen's d, the standard measure in behavioral science:

$$d = (\text{Treatment Mean} - \text{Control Mean}) / \text{Pooled Standard Deviation}$$

When you specify $d = 0.5$, the system adjusts response distributions so that the mean difference between conditions matches your target. This is achieved through:

1. **Semantic parsing** of condition names to determine effect direction
2. **Graduated adjustments** applied at the individual response level
3. **Validation checks** confirming achieved effects match targets

Scale Reliability Modeling

Multi-item scales exhibit realistic internal consistency (Cronbach's alpha) through a factor model approach:

$$\text{Response} = \lambda * \text{Common_Factor} + \sqrt{1 - \lambda^2} * \text{Unique_Error}$$

Where λ (factor loading) is derived from the target reliability. Items measuring the same construct share common variance while retaining item-specific variation, producing alpha values typically ranging 0.75-0.90.

Response Style Modeling

The system models several well-documented response styles:

Acquiescence Bias: Tendency to agree with statements, producing higher means on positively-worded items. Modeled as a per-participant offset.

Extreme Response Style: Tendency to use scale endpoints. Higher extremity = more responses at 1 or 7 (on 7-point scales).

Social Desirability: Inflation of socially favorable responses. Applied proportionally based on item content.

Midpoint Avoidance: Cultural variation in willingness to use neutral midpoint. East Asian samples typically show lower midpoint avoidance than Western samples.

Survey Flow Logic

The system respects your experimental design by tracking which questions each participant would actually see:

- **Block-level conditions:** Participants only receive responses for their assigned condition's blocks
- **Display logic:** Questions with condition-specific visibility rules are handled appropriately
- **Factorial designs:** Crossed conditions (e.g., AI x Hedonic, AI x Utilitarian) are properly parsed

Open-Ended Response Generation

Open-ended text responses are generated using a two-tier system that maximizes realism and uniqueness.

Tier 1: AI-Powered Generation (Primary)

When available, responses are generated by a large language model (LLM) that receives the full experimental context — study description, condition assignment, and participant persona — and produces natural, question-specific text that mirrors real survey responses.

Zero-configuration AI: The tool ships with built-in API keys for three free LLM providers, so AI-powered responses work out of the box with no setup required:

Provider	Model	Free Tier
Groq (primary)	Llama 3.3 70B Versatile	14,400 requests/day
Cerebras (failover)	Llama 3.3 70B	1M tokens/day
OpenRouter (failover)	Mistral Small 3.1 24B	Free model tier

If one provider reaches its rate limit, the system automatically tries the next. Users can optionally provide their own free API key from any of these providers for additional capacity.

Key features:

- Batch generation:** 20 persona-guided responses are generated per API call, each tailored to a different participant profile (varying in verbosity, formality, engagement level, and sentiment)
- Draw-with-replacement pooling:** A pool of LLM-generated base responses is pre-built for each question × condition × sentiment bucket; individual participants draw from this pool with deep persona-driven variation applied, ensuring no two responses are identical even when they share a common base
- 7-layer deep variation:** Each drawn response passes through word-level micro-variation, sentence restructuring, verbosity control, formality adjustment, engagement modulation, typo injection, and synonym substitution — producing unique output for every participant
- Smart pool scaling:** Pool size automatically adapts to sample size (using $\sqrt{n} \times 3 + 10$, clamped to [30, 80] per bucket), balancing API efficiency with response diversity
- 3-provider failover chain:** Groq → Cerebras → OpenRouter → user's own key, ensuring maximum uptime with no single point of failure

Tier 2: Template-Based Generation (Fallback)

If the AI service is unavailable, the system falls back to a comprehensive template engine covering 225+ research domains and 40 question types. This ensures:

- Uniqueness:** No two participants give identical responses
- Context-awareness:** Responses reference the experimental manipulation when appropriate
- Condition-specificity:** Only participants who would see a question receive a response
- Length variation:** Verbosity varies by persona and question type

Example Responses

For a question "What did you think about the AI recommendations?":

Treatment participant (saw AI):

"I found the AI-generated recommendations to be helpful. The system seemed to capture my preferences well. I appreciated how it addressed my needs effectively."

Control participant (no AI):

(Empty - this participant didn't see this question)

Research Domain Coverage: 225+ Scientific Areas

The response generation system has been trained on **hundreds of scientific insights** drawn from decades of research across the social and behavioral sciences. This extensive knowledge base enables the tool to generate contextually appropriate responses for virtually any research topic you might study.

Major Research Fields

The system covers **33 major research categories** with over **225 specialized domains**:

Field	Domains	Example Topics
Behavioral Economics	12	Trust games, dictator games, ultimatum games, public goods, risk preferences, time preferences, loss aversion, framing effects, anchoring, sunk cost fallacy
Social Psychology	15	Intergroup relations, social identity, norms, conformity, prosocial behavior, cooperation, fairness, social influence, attribution, stereotypes, prejudice, empathy
Political Science	10	Polarization, partisanship, voting behavior, media effects, policy attitudes, civic engagement, political trust, ideology, misinformation
Consumer/Marketing	10	Brand perception, advertising effectiveness, purchase intent, brand loyalty, price perception, service quality, customer satisfaction, word-of-mouth
Organizational Behavior	10	Leadership, teamwork, motivation, job satisfaction, organizational commitment, work-life balance, employee engagement, organizational culture
Technology & AI	10	AI attitudes, privacy concerns, automation, algorithm aversion, technology adoption, social media, digital wellbeing, human-AI interaction
Health Psychology	10	Medical decision-making, wellbeing, health behaviors, mental health, vaccination attitudes, pain management, patient-provider communication

Ethics & Moral Psychology	8	Moral judgment, ethical dilemmas, moral emotions, values, ethical leadership, corporate ethics, research ethics
Environmental Psychology	8	Sustainability, climate attitudes, pro-environmental behavior, green consumption, conservation, energy behavior
Cognitive Psychology	8	Decision-making, memory, attention, reasoning, problem-solving, cognitive biases, metacognition

Additional Specialized Domains

The system also covers:

- **Education** (8 domains): Learning, academic motivation, teaching effectiveness, online learning, educational technology
- **Developmental Psychology** (6 domains): Parenting, childhood development, aging, life transitions
- **Clinical Psychology** (6 domains): Anxiety, depression, coping strategies, therapy attitudes, stress
- **Communication** (6 domains): Persuasion, media effects, interpersonal communication, narrative processing
- **Neuroeconomics** (6 domains): Reward processing, impulse control, emotional regulation, cognitive load
- **Sports Psychology** (6 domains): Athletic motivation, team dynamics, performance anxiety, fan behavior
- **Legal Psychology** (6 domains): Jury decision-making, witness memory, procedural justice
- **Food Psychology** (6 domains): Eating behavior, food choice, nutrition knowledge, body image
- **Human Factors** (6 domains): User experience, interface design, safety behavior, human error
- **Cross-Cultural** (5 domains): Cultural values, acculturation, cultural identity
- **Positive Psychology** (5 domains): Gratitude, resilience, flourishing, life satisfaction
- **Financial Psychology** (6 domains): Financial literacy, investment behavior, retirement planning
- **Personality Psychology** (6 domains): Big Five traits, narcissism, dark triad, self-concept
- **Social Media Research** (6 domains): Online identity, digital communication, influencer effects

Topic-Specific Response Generation

For each research domain, the system maintains specialized knowledge about:

1. **Key constructs and terminology** used in that field
2. **Typical response patterns** observed in empirical studies
3. **Common participant concerns and attitudes** documented in the literature
4. **Domain-specific language and phrasing** that real participants use

This means when you study cooperation in public goods games, the system generates responses that reference concepts like "contribution," "free-riding," "collective benefit," and "reciprocity"—just as real participants would. Similarly, for AI trust research, responses naturally mention "algorithmic recommendations," "automation," "reliability," and "transparency."

Scientific Foundation

The domain knowledge is built on insights from:

- **Classic experiments:** Kahneman & Tversky's prospect theory work, Milgram's obedience studies, Asch's conformity experiments
 - **Meta-analyses:** Aggregated findings from hundreds of studies in each domain
 - **Replication projects:** Many Labs, Psychological Science Accelerator, and other large-scale replication efforts
 - **Contemporary research:** Recent publications in top journals (JPSP, Psychological Science, JEP:G, Management Science, etc.)
-

Data Quality Features

Attention Check Simulation

Configurable proportion of participants fail attention checks, matching real-world rates. Failed attention checks are flagged for potential exclusion.

Careless Response Detection

The system can identify (and optionally flag or exclude) simulated careless responses:

- **Straight-lining:** Same response repeated across items
- **Alternating patterns:** Systematic alternation (1-7-1-7)
- **Midpoint overuse:** Excessive neutral responses
- **Response time anomalies:** Unrealistically fast completion

Validation Metrics

Generated datasets include quality metrics:

- Achieved effect sizes with confidence intervals
 - Condition balance verification
 - Missing data rates
 - Response distribution statistics
-

Technical Specifications

Input Requirements

- **Option A:** Qualtrics Survey Format (.qsf) file — for researchers with existing surveys
- **Option B:** Plain-language experiment description — for researchers at the design stage
- Internet connection for web interface

Output Format

- **CSV file** compatible with R, SPSS, Stata, Python
- **Instructor report** (HTML) with comprehensive statistical analyses, persona breakdowns, effect size verification, trait profiles by condition, and visualization
- **Metadata** JSON with simulation parameters
- **Analysis scripts** auto-generated for R, Python, SPSS, and Stata

Supported Question Types

Type	Example	Input Methods
Likert Scales	7-point agreement scales	QSF, Builder
Sliders	Visual analog scales (0-100)	QSF, Builder
Matrix Tables	Multi-item scales with shared options	QSF, Builder
Multiple Choice	Single selection questions	QSF
Text Entry	Open-ended responses	QSF, Builder
Numeric Input	Willingness to pay, quantities	QSF, Builder
Semantic Differential	Bipolar adjective scales	QSF, Builder
Binary	Yes/No, True/False	QSF, Builder
Rank Order	Preference rankings	QSF
Heatmaps	Click coordinate data	QSF

Supported Experimental Designs

Design	Example	Detection
Between-subjects	Treatment vs Control	Automatic
Factorial (2x2)	AI x Product Type	Automatic (NxM)
Factorial (3x2)	Annotation x Product	Automatic
Factorial (NxM)	Any crossed design	Automatic
Multi-level	Low / Medium / High	Automatic
Mixed designs	Between + within factors	QSF path
Complex branching	Skip logic, visibility rules	QSF path

The conversational builder automatically detects factorial designs from natural language input and generates all crossed conditions. For example, entering "3 (Source: AI vs Human vs None) x 2 (Product: Hedonic vs Utilitarian)" produces 6 conditions with proper x notation.

Frequently Asked Questions

Is this generating "fake data"?

No—this is **synthetic data generation**, a legitimate research methodology. The tool produces data with known statistical properties for specific purposes:

- Testing analysis code before real data collection
- Teaching data analysis with realistic datasets
- Power analysis and sample size planning
- Developing preprocessing pipelines

Synthetic data should never be misrepresented as real participant data in publications.

How realistic are the responses?

The responses exhibit statistical properties matching published research on human survey behavior:

- Mean responses around 5.0-5.5 on 7-point scales (documented positive response bias)
- Standard deviations of 1.2-1.8 (typical for Likert data)
- Cronbach's alphas of 0.75-0.90 for multi-item scales
- Effect sizes within +/-0.15 of specified targets

Can I use this for any survey?

The tool is optimized for behavioral science experiments with:

- Clear experimental conditions (treatment/control)
- Likert or similar scale DVs
- Defined effect size expectations

It may not be suitable for purely exploratory surveys or complex longitudinal designs.

Do I need a Qualtrics survey file?

No. As of version 1.3, you can describe your experiment in plain language using the **Conversational Builder**. The system parses your natural language description to extract conditions, scales, and open-ended questions. This is especially useful for:

- Early-stage study design before building the actual survey
- Quick pilot data generation
- Teaching contexts where students describe hypothetical experiments
- Power analysis before IRB submission

What factorial designs are supported?

The system supports any N×M crossed factorial design:

- **2×2**: "AI (present, absent) × Trust (high, low)"
- **3×2**: "3 (Source: AI vs Human vs None) × 2 (Product: Hedonic vs Utilitarian)"
- **2×2×2**: Three-factor designs with automatic crossing
- **Custom**: Any combination using "vs", commas, or numbered lists

The system automatically generates all crossed conditions and displays them in a design table.

How are effect directions determined?

The system parses condition names to determine which should produce higher/lower responses:

- **Positive indicators:** "high", "treatment", "reward", "positive"
- **Negative indicators:** "low", "control", "loss", "negative"

For complex designs, effect direction can be specified manually in the design review.

What does the instructor report contain?

The comprehensive HTML report includes:

- **Study overview:** Design, conditions, factors, scales, sample size
 - **Research context:** Domain, input method, participant characteristics, persona domains
 - **Statistical analysis:** Per-DV descriptive statistics, ANOVA/t-tests, effect sizes, visualizations
 - **Persona analysis:** Distribution table, per-condition persona counts, personality trait profiles
 - **Effect size verification:** Configured vs. observed effects with Cohen's d interpretation
 - **Data quality:** Exclusion breakdown (speed, attention, straight-lining), validation corrections
 - **Categorical analysis:** Condition × Gender cross-tabulation with chi-squared test
 - **Executive summary:** AI-generated synthesis of key findings
 - **Scientific references:** Full citations for the methodological foundations
-

Getting Started

Option A: With a Qualtrics Survey

1. **Export your Qualtrics survey** as a .qsf file (Survey > Tools > Import/Export > Export Survey)
2. **Access the tool** at the provided URL
3. **Upload your .qsf file** and review the detected structure
4. **Configure parameters** for your specific needs
5. **Generate and download** your synthetic dataset

Option B: With the Conversational Builder

1. **Access the tool** and select "Describe my study" in the Study Input tab
 2. **Enter your study title** and a brief description
 3. **Describe your conditions** (the system detects simple, factorial, and multi-level designs)
 4. **Describe your scales/DVs** (supports standard abbreviations, paragraph format, and detailed specs)
 5. **Add open-ended questions** if applicable
 6. **Review and adjust** the auto-detected design in the Design tab
 7. **Generate and download** your synthetic dataset
-

Scientific References

The simulation algorithms are grounded in established survey methodology research:

1. **Cohen, J. (1988).** Statistical power analysis for the behavioral sciences. *Effect size conventions and calculations*.
 2. **Krosnick, J. A. (1991).** Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236. *Optimizing vs. satisficing response strategies*.
 3. **Greenleaf, E. A. (1992).** Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351. *Extreme response style measurement and prevalence*.
 4. **Billiet, J. B., & McClendon, M. J. (2000).** Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628. *Acquiescence bias quantification*.
 5. **Meade, A. W., & Craig, S. B. (2012).** Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455. *Careless response detection methods*.
 6. **Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003).** One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331-363. *Meta-analytic effect size distributions*.
-

Citation

If you use this tool in your research or teaching, please acknowledge:

Dimant, E. (2025). Behavioral Experiment Simulation Tool (Version 1.4.10) [Computer software].

Changelog Highlights

Version 1.4 (Latest)

- **AI-powered open-ended responses:** LLM-generated text with draw-with-replacement pooling and 7-layer deep persona variation
- **Multi-provider failover:** Groq, Cerebras, and OpenRouter with automatic key detection
- **Smart pool scaling:** Automatically adapts response pool size to study sample size
- **Template fallback:** Seamless degradation to template engine when AI is unavailable

Version 1.3

- **Conversational Builder:** Describe experiments in plain language — no QSF file required
- **Automatic factorial detection:** Parses NxM designs from natural language (e.g., "3 × 2, between-subjects")
- **Comprehensive instructor report:** Persona distribution tables, personality trait profiles by condition, effect size verification, exclusion breakdowns
- **Scale auto-detection:** Recognizes detailed academic scale formats, validated instruments (BFI-10, PANAS, etc.), numeric inputs, binary measures
- **Custom persona weights:** Adjust response style distributions for domain-specific realism
- **Domain-specific personas:** 225+ research domains influence which persona archetypes are activated

Version 1.2

- Enhanced persona system with 50+ behavioral archetypes across 15 research domains
- Factorial design tables with visual cell numbering

- Effect size specification with Cohen's d calibration
- Auto-generated analysis scripts for R, Python, SPSS, and Stata

Version 1.0

- Initial release with QSF upload, persona-based response generation, and basic instructor reports
-

Support and Contact

For questions, feature requests, or collaboration inquiries, please contact through official university channels.

Version 1.4.10 | Proprietary Software | All Rights Reserved

Developed by Dr. Eugen Dimant