

# 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
- **Scale type auto-correction:** Single-item DVs are automatically identified (not mislabeled as "Likert Scale"); multi-item scales are properly categorized by type (matrix, numbered items, single item). Scale min/max values are propagated from QSF detection.
- Add or remove measures as needed
- **Custom demographic variables:** Add demographic questions beyond the defaults (Age, Gender). Quick-add templates include Political Orientation, Education Level, Ethnicity, Household Income, Employment Status, Religion, and Party Identification. Each demographic is fully customizable:

- **Categorical:** Edit options and their probability weights (e.g., Political Orientation with "Very Conservative" through "Very Progressive")
- **Ordinal:** Set ordered categories with center-weighted distribution
- **Numeric:** Configure mean, standard deviation, and min/max bounds (e.g., household income)
- **Distribution preview:** See the expected distribution before generating
- 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 (including custom variables) 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, sunk cost, anchoring
- **Social Psychology:** Social identity, group dynamics, conformity, prosocial behavior, intergroup relations, attitudes, persuasion, social comparison
- **Cognitive Psychology:** Decision-making heuristics, cognitive biases, memory, attention, reasoning processes, construal level
- **Organizational Behavior:** Leadership, motivation, job satisfaction, team dynamics, workplace attitudes, power dynamics
- **Political Psychology:** Polarization, partisanship, civic engagement, media effects, political trust, sacred values
- **Consumer Behavior:** Brand perception, purchase intent, product evaluation, advertising effectiveness, choice architecture
- **Moral Psychology:** Ethical judgment, values, moral emotions, fairness perceptions, moral cleansing, sacred value tradeoffs
- **Health Psychology:** Medical decisions, wellbeing, health behaviors, treatment preferences, gratitude interventions
- **Narrative & Communication:** Narrative transportation, story persuasion, source credibility, elaboration likelihood, inoculation theory
- **Digital & Technology:** Attention economy, phone distraction, social media comparison, digital wellbeing, algorithm aversion
- **Positive Psychology:** Gratitude interventions, savoring, best possible self, acts of kindness, growth mindset

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  $\lambda$  (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, with **domain-sensitive intensity**: highly sensitive topics (prejudice, dishonesty) receive 1.5x the social desirability adjustment,

while factual/behavioral reports receive only 0.5x. Based on Nederhof (1985) and Paulhus (2002).

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

## **Persona-Demographic Coupling**

Custom demographic variables are not assigned randomly—they are **coupled to each participant's behavioral persona** to create realistic correlations between demographic characteristics and response patterns, as observed in real survey data.

The coupling algorithm uses a **swap-sort** approach that preserves exact marginal distributions while creating realistic associations:

1. **Affinity lookup:** A mapping defines which persona types tend toward which demographic values (e.g., "Extreme Responder" personas are more likely to hold strong political orientations; "Engaged" personas correlate with higher education levels)
2. **Trait-based soft coupling:** Beyond explicit persona-demographic mappings, participant traits like extremity, attention, and consistency create additional correlations (e.g., high-extremity participants skew toward extreme political positions)
3. **Distribution-preserving swaps:** Rather than generating new values, the algorithm only swaps existing demographic assignments between participants when a swap improves the persona-demographic fit—ensuring the overall distribution exactly matches the user's specified weights

This means a simulated dataset where 35% of participants are "Engaged Responders" will show a natural correlation between engagement and education level, mirroring real-world patterns without distorting the researcher's target demographic distributions.

## **Behavioral Coherence (Rating–Text Consistency)**

A key advancement is the **behavioral coherence pipeline** that ensures each simulated participant's numeric ratings and open-text responses tell a coherent story. The same person who rates trust at 6-7/7 writes positively about trust; a participant who straight-lines 4s across all items writes brief, disengaged text.

This is enforced through:

1. **Behavioral profiling:** Each participant's numeric pattern (mean, variability, straight-lining) is computed before text generation
2. **Profile-guided text:** The behavioral profile flows to all text generators, constraining tone, length, and engagement
3. **Post-generation validation:** Text responses are checked against numeric patterns and corrected if mismatched
4. **Cross-item consistency tracking:** Participants who fail reverse-coded items are more likely to fail subsequent ones, matching Woods (2006) findings that reversal failure is trait-like within session

## **Reverse-Coded Item Modeling**

Reverse-coded items receive sophisticated handling that goes beyond simple scale inversion:

- **Engagement-dependent accuracy:** Engaged respondents correctly reverse ~95% of the time; careless respondents only ~30-50%
- **Acquiescence interaction:** Even respondents who correctly reverse show partial acquiescence pull (~0.5 point, Weijters et al. 2010)
- **Cross-item failure consistency:** A participant who fails one reverse item is more likely to fail the next (trait-like within session)

## **Response Validation Layer**

Generated responses are validated against expected patterns for each persona type:

- **Longstring detection:** Flags unrealistic straight-lining for engaged personas
- **IRV checks:** Ensures response variability matches persona engagement level
- **Endpoint utilization:** Verifies extreme response style personas actually use scale endpoints

## **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:**

1. **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)
2. **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
3. **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
4. **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

5. **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. As of v1.0.8.0, the template system uses a **compositional architecture** that produces highly varied, topic-grounded responses:

1. **Intent-driven composition:** Each response is assembled from opener + intent-matched core + domain-enriched elaboration + coda. Question intent is classified into 8 categories (opinion, explanation, description, emotional reaction, evaluation, prediction, causal explanation, decision explanation) and templates are selected accordingly
2. **40+ domain vocabulary sets:** Specialized terminology for clinical/mental health, sports, legal, food, developmental, personality, cognitive, neuroscience, financial, cross-cultural, and 30+ more domains ensures responses use field-appropriate language
3. **Rich question-text mining:** 33 action verb patterns, 24 object/target pattern groups, and 15 key phrase patterns extract the actual topic from the question text for template insertion
4. **Domain-gated condition modifiers:** Condition-specific personalizations (e.g., "As someone who leans progressive") are only applied when the domain matches — political modifiers only fire for political studies, health modifiers only for health studies
5. **Behavioral coherence:** Templates are post-processed to match the participant's numeric response pattern — straight-liners get truncated text, extreme raters get intensified language, high social desirability personas get qualifying hedges
6. **25 careless response templates:** Even low-effort responses reference the actual topic ("trump is ok i guess") rather than generic off-topic text ("fine")
7. **Context-awareness:** Responses reference the experimental manipulation when appropriate
8. **Condition-specificity:** Only participants who would see a question receive a response

## LLM Exhaustion Recovery

When free AI providers reach their rate limits during generation, the system provides a **transparent, user-controlled recovery flow** rather than silently switching methods:

1. **Immediate notification:** The user sees a clear message showing how many participants were completed with AI and how many remain
2. **Four simultaneous options** presented as equal choices:
  - **Retry Built-in AI:** Wait for rate limits to reset and try again
  - **Use Your Own API Key:** Enter a personal (free) key from Groq, Cerebras, or OpenRouter for unlimited capacity
  - **Template Engine:** Complete remaining participants with the 225-domain template system
  - **Adaptive Behavioral Engine:** Switch to the paradigm-aware behavioral simulation engine
3. **Data preservation:** Responses already generated by AI are preserved; only remaining participants use the selected fallback method
4. **Source transparency:** The final dataset includes metadata showing which responses were AI-generated vs. template-generated

This ensures the researcher always knows—and controls—the data source for their simulated dataset.

## Self-Healing Error Pipeline

The system includes an **automatic error logging and self-healing pipeline** that captures generation failures and facilitates rapid fixes:

- 1. Automatic error capture:** Every generation error is logged with full context—traceback, session state snapshot, generation method, timing, and phase
- 2. Error fingerprinting:** Errors are deduplicated using a SHA-256 fingerprint of the error type + last traceback frame, so recurring issues are counted rather than duplicated
- 3. Structured reporting:** On software updates (version changes), the system automatically generates a structured PENDING\_FIXES.md report listing all unresolved errors with their full context
- 4. Privacy-preserving:** Sensitive keys (API keys, passwords, binary content) are automatically redacted from error logs
- 5. Fix tracking:** Each error carries a status (pending → acknowledged → fixed) with the version that resolved it

This pipeline ensures that user-encountered errors are never lost and are systematically addressed in subsequent updates.

## 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

<b>Political Science</b>	10	Polarization, partisanship, voting behavior, media effects, policy attitudes, civic engagement, political trust, ideology, misinformation
<b>Consumer/Marketing</b>	10	Brand perception, advertising effectiveness, purchase intent, brand loyalty, price perception, service quality, customer satisfaction, word-of-mouth
<b>Organizational Behavior</b>	10	Leadership, teamwork, motivation, job satisfaction, organizational commitment, work-life balance, employee engagement, organizational culture
<b>Technology &amp; AI</b>	10	AI attitudes, privacy concerns, automation, algorithm aversion, technology adoption, social media, digital wellbeing, human-AI interaction
<b>Health Psychology</b>	10	Medical decision-making, wellbeing, health behaviors, mental health, vaccination attitudes, pain management, patient-provider communication
<b>Ethics &amp; Moral Psychology</b>	10	Moral judgment, ethical dilemmas, moral emotions, values, ethical leadership, corporate ethics, moral cleansing, sacred values, moral licensing
<b>Environmental Psychology</b>	8	Sustainability, climate attitudes, pro-environmental behavior, green consumption, conservation, energy behavior
<b>Cognitive Psychology</b>	8	Decision-making, memory, attention, reasoning, problem-solving, cognitive biases, metacognition
<b>Narrative &amp; Communication</b>	8	Narrative transportation, story persuasion, source credibility, elaboration likelihood, inoculation, message framing

Digital & Attention	6	Phone distraction, notification effects, media multitasking, digital detox, screen time, social media comparison
Positive Psychology	8	Gratitude, savoring, kindness interventions, best possible self, growth mindset, resilience, flourishing

## 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 NxM crossed factorial design:

- **2x2:** "AI (present, absent) × Trust (high, low)"
- **3x2:** "3 (Source: AI vs Human vs None) × 2 (Product: Hedonic vs Utilitarian)"
- **2x2x2:** 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 x 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:

### Core Survey Methodology

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.
3. **Greenleaf, E. A. (1992).** Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
4. **Billiet, J. B., & McClelland, M. J. (2000).** Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628.
5. **Meade, A. W., & Craig, S. B. (2012).** Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455.

6. **Paulhus, D. L. (2002)**. Socially desirable responding. *Journal of Personality Assessment*, 40, 13-44.
7. **Nederhof, A. J. (1985)**. Methods of coping with social desirability bias. *European Journal of Social Psychology*, 15, 263-280.
8. **Woods, C. M. (2006)**. Careless responding to reverse-worded items. *Journal of Psychoeducational Assessment*, 24, 207-220.
9. **Weijters, B., et al. (2010)**. The effect of rating scale format on response styles. *International Journal of Research in Marketing*, 27, 236-247.

## **Behavioral Economics & Game Theory**

10. **Engel, C. (2011)**. Dictator games: A meta study. *Experimental Economics*, 14, 583-610.
11. **Dimant, E. (2024)**. Partisan intergroup discrimination in economic games.
12. **Iyengar, S., & Westwood, S. J. (2015)**. Fear and loathing across party lines. *American Journal of Political Science*, 59, 690-707.
13. **Kahneman, D., & Tversky, A. (1979)**. Prospect theory. *Econometrica*, 47, 263-291.

## **Paradigms & Phenomena**

14. **Green, M. C., & Brock, T. C. (2000)**. The role of transportation in the persuasiveness of public narratives. *JPSP*, 79, 701-721.
  15. **Festinger, L. (1954)**. A theory of social comparison processes. *Human Relations*, 7, 117-140.
  16. **Emmons, R. A., & McCullough, M. E. (2003)**. Counting blessings versus burdens. *JPSP*, 84, 377-389.
  17. **Zhong, C. B., & Liljenquist, K. (2006)**. Washing away your sins: Threatened morality and physical cleansing. *Science*, 313, 1451-1452.
  18. **Ward, A. F., et al. (2017)**. Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *JACR*, 2, 140-154.
  19. **Tetlock, P. E., et al. (2000)**. The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *JPSP*, 78, 853-870.
  20. **Podsakoff, P. M., et al. (2003)**. Common method biases in behavioral research. *JAP*, 88, 879-903.
- 

## **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 1.2.0.5 (Latest)**

- **Custom demographic variables**: Full flexibility to add and customize demographic questions (Political Orientation, Education, Ethnicity, Income, Employment, Religion, Party ID) with editable options, weights, and distributions
- **Persona-demographic coupling**: Swap-sort algorithm creates realistic correlations between persona types and demographic values while preserving exact marginal distributions
- **Scale type auto-correction**: Single-item DVs properly identified; scale type/min/max propagated from QSF detection
- **LLM exhaustion recovery**: Transparent 4-option UI when free AI providers are exhausted—user always controls the data source

- **Self-healing error pipeline:** Automatic error logging with full context capture, fingerprinting, deduplication, and structured fix reporting on software updates
- **Stale-phase recovery:** Unprotected code paths in the generation pipeline now wrapped in comprehensive error handling

### **Version 1.0.4.9**

- **5 new research paradigm domains:** Narrative transportation, social comparison, gratitude interventions, moral cleansing/sacred values, digital attention economy
- **6 new domain-specific personas:** Narrative Thinker, Social Comparer, Grateful Optimist, Moral Absolutist, Digital Native, Financial Deliberator
- **Enhanced social desirability:** Domain-sensitive construct detection for moral identity, gratitude, digital habits, social comparison
- **Cross-item reverse-failure tracking:** Participants who fail one reverse item are more likely to fail subsequent ones (Woods 2006)
- **Response validation layer:** Post-generation checks for longstring, IRV, and endpoint utilization anomalies
- **Expanded domain templates:** 5 new domain template sets with both explanation and evaluation question types
- **Behavioral coherence pipeline:** Rating-text consistency ensures numeric patterns match open-text tone
- **Front-facing methods documentation:** Comprehensive update with new paradigms and scientific references

### **Version 1.0.4.8**

- **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 x 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 1.2.0.5 | Proprietary Software | All Rights Reserved*

*Developed by Dr. Eugen Dimant*