# Blind Self-Recognition Experiment

*Product and Research Design Document*

*The map is not the territory, but neither is*

*the territory obliged to ignore the map.*

# For Science

*Toby Godden*

Working title: Which Sky Is Yours?

Project: Chronometer / Geometric Symbolic Reasoning

Author: Toby Godden

Status: Experimental design, version 0.1

□

1. Purpose

This experiment tests whether an astrological interpretation generated from a participant’s actual birth geometry is recognised as more personally accurate than interpretations generated from other plausible birth geometries.

Each participant supplies:

* date of birth;
* exact or approximate time of birth;
* place of birth.

The system computes one interpretation from the participant’s actual birth data and ten interpretations from matched control birth records.

The eleven interpretations are shuffled and presented without dates, charts, planet names or any indication of which is genuine.

The participant answers one central question:

Which of these descriptions is most like you?

The system records whether the participant selected the interpretation generated from their actual birth geometry.

Across many participants, the experiment measures whether people identify their own interpretation more often than chance.

With eleven choices, random selection would identify the correct interpretation approximately 1 time in 11, or 9.09% of the time.

## 2. Research question

### Primary question

Do participants identify the interpretation generated from their actual birth geometry more frequently than the chance rate of 1 in 11?

### Primary hypothesis

Participants will select their genuine interpretation at a rate greater than 9.09%.

### Null hypothesis

Participants will select their genuine interpretation at approximately the chance rate of 9.09%.

### Secondary questions

The experiment may also investigate:

* whether the genuine interpretation receives a higher rating than the ten controls;
* whether the genuine interpretation appears within the participant’s top three choices;
* whether confidence predicts accuracy;
* whether birth-time precision affects accuracy;
* whether experienced astrologers perform differently from participants with no astrological knowledge;
* whether particular graph features, such as tight aspects or angularity, contribute more strongly to successful identification;
* whether the result changes when explicit astrological terminology is removed.

□

3. Core participant experience

The experiment should feel like a personality-recognition exercise rather than an astrology quiz.

Participant journey

1. The participant creates an account or enters through a unique participation link.
2. They read the consent and privacy information.
3. They enter their birth date, birth time and birthplace.
4. They indicate the source and reliability of their birth time.
5. The system generates eleven interpretations.
6. The interpretations are shown anonymously and in random order.
7. The participant reads and rates each interpretation.
8. The participant selects the one they believe describes them best.
9. They optionally choose a second and third best match.
10. They report their confidence.
11. The response is permanently recorded.
12. The participant sees either:

* an immediate result;
* a delayed result after the study closes;
* or no individual reveal, depending on the selected research protocol.

The recommended research version delays the reveal until the participant has completed every question and the submission has been locked.

□

4. The eleven interpretations

Each participant receives:

* 1 target interpretation, generated from their actual birth data;
* 10 control interpretations, generated from genuine but non-target birth geometries.

The controls should appear every bit as plausible and internally coherent as the target.

Why controls should not be completely arbitrary

Random dates, times and locations could accidentally create detectable differences between the target and controls.

For example:

* the target may use a UK location while controls span implausibly distant latitudes;
* control charts may contain different aspect densities;
* some interpretations may be much longer because their charts contain more selected relations;
* seasonal or generational placements may make one interpretation visibly different;
* impossible or statistically unusual combinations could reveal which graph came from a real birth record.

Therefore, the controls should be random within matching constraints.

Recommended control generation

For each participant, generate ten control birth records matched approximately on:

* birth-year range;
* season or month;
* latitude band;
* time-of-day band;
* birth-time precision;
* chart complexity or selected relation count.

The controls may come from one of two sources.

Preferred method: real participant controls

Use birth records belonging to ten other consented participants.

Advantages:

* every control is an astronomically genuine chart;
* real birth-date distributions are preserved;
* impossible synthetic combinations cannot occur;
* controls can be reused through a balanced assignment system.

No participant identity or biography is shared between users.

Initial prototype method: generated control moments

Generate valid dates, times and terrestrial locations according to preregistered matching rules.

These are not random edges. They are complete, astronomically possible moments passed through the same Chronometer engine.

For example, a control generator might:

* retain the participant’s year within ±2 years;
* retain the calendar month or season;
* sample a valid time within the same six-hour band;
* sample a real settlement within the same latitude band;
* reject graphs outside the target range of aspect density.

The generation rules must be fixed before analysing results.

□

5. Interpretation generation

All eleven interpretations must be generated using exactly the same pipeline.

Fixed generation pipeline

1. Compute astronomical positions.
2. Construct the relational graph.
3. Identify the permitted aspect set and weights.
4. Bind the frozen symbolic lexicon to the graph.
5. pass the annotated graph to the language model.
6. Validate all cited entities and relations.
7. Render the final interpretation using a fixed structure and word count.

Information supplied to the model

The model may receive:

* body identifiers;
* signs, where included in the registered test;
* measured aspects;
* exact orbs;
* applying or separating status;
* angularity;
* permitted symbolic definitions;
* neutral writing instructions.

The model must not receive:

* participant name;
* biography;
* occupation;
* gender, unless gender is explicitly part of a preregistered study;
* interests;
* previous chat history;
* account data;
* which graph is the target;

* any indication that a participant will be judging the output.

Version locking

The following must be recorded for every generated interpretation:

* Chronometer engine version;
* symbolic lexicon version;
* prompt version;
* model name and exact version where available;
* generation settings;
* validator version;
* graph hash;
* interpretation hash;
* generation timestamp.

Once a study begins, these components should not be changed without starting a new experimental cohort.

□

6. Interpretation format

Every interpretation should have the same visible structure.

Suggested structure:

Central character

A concise description of the person’s central orientation or recurring mode of engagement.

Inner tensions

Two or three tensions or contrasting tendencies suggested by the graph.

Relationships

A description of how the person may approach intimacy, collaboration or interpersonal boundaries.

Work and expression

A description of likely approaches to work, creativity, responsibility or public action.

Change and pressure

A description of how the person may respond to conflict, transition, uncertainty or constraint.

Summary

A short synthesis of the whole reading.

Each interpretation should:

* contain the same number of sections;
* remain within a narrow word-count range;
* avoid dates, planet names and astrological jargon;
* avoid gendered assumptions;
* avoid highly generic flattery where possible;
* contain both strengths and tensions;
* avoid medical, diagnostic, criminal or traumatic claims;
* not predict specific events;

* not reveal geographical or generational clues.

A suitable initial length is approximately 300–450 words per interpretation.

Eleven readings of that length create a significant reading burden. A shorter first-stage format of 120–180 words may produce better completion rates.

A later study could compare short and long interpretations.

□

7. Participant interface

Screen 1: Introduction

Headline:

Eleven descriptions. One was generated from your birth geometry.

Supporting text:

Read each description carefully and choose the one that feels most specifically like you. The remaining descriptions were generated using other astronomically possible birth moments. The system does not know anything about your life beyond the birth information you provide.

Do not reveal the expected hypothesis or discuss astrology’s validity before the task.

Screen 2: Consent

The participant confirms that:

* they are at least the minimum permitted age;
* they understand this is a research experiment;
* participation is voluntary;
* they may withdraw before submitting;
* birth data will be processed for the experiment;
* generated descriptions are not medical or psychological assessments;
* anonymised results may be used in publications;

* they understand the retention and deletion policy.

Screen 3: Birth information

Fields:

* birth date;
* birth time;
* “I do not know my exact birth time” option;
* birthplace search;
* resolved latitude and longitude;
* timezone at birth;
* source of birth time.

Birth-time source options:

* official birth certificate or registration;
* hospital or family record;
* parent or relative;
* personal memory;
* rectified by an astrologer;
* approximate or unknown.

Confidence options:

* exact to the minute;
* within 5 minutes;
* within 15 minutes;
* within 30 minutes;
* within 1 hour;
* approximate;
* unknown.

Participants with unknown times may be excluded from the primary analysis but included in a separate exploratory cohort.

Screen 4: Generation

The system generates the target and controls without exposing their provenance.

The participant sees a neutral progress message:

Preparing the eleven descriptions.

Screen 5: Reading and rating

Each interpretation appears as a numbered card:

* Description A
* Description B
* Description C
* and so forth.

The labels are randomly assigned for every participant.

For each description, the participant provides:

Overall fit

* 1 — Not like me
* 2 — Slightly like me
* 3 — Partly like me
* 4 — Mostly like me
* 5 — Very strongly like me

Optional additional ratings:

* specificity;
* emotional resonance;
* accuracy of tensions;
* accuracy of relationship description;
* accuracy of work or creative description.

The participant must rate every description before moving to final selection.

The cards should be shown one at a time or in a controlled reading sequence. Displaying all eleven side by side would encourage superficial comparison and create major mobile usability problems.

Screen 6: Final selection

Prompt:

Which single description is most like you?

The participant chooses one.

Then:

Which two descriptions would you place second and third?

Finally:

How confident are you that your first choice is the description generated from your actual birth information?

Confidence scale:

* pure guess;
* slightly confident;
* moderately confident;
* very confident;
* almost certain.

The participant may also answer:

Did any description seem obviously different from the others?

This helps detect formatting or generation leakage.

Screen 7: Submission lock

Before submission:

Your answers cannot be changed after submission. This prevents results being altered after the correct description is revealed.

The system records the submission and seals it.

Screen 8: Result

For an informal public prototype, show:

* whether the participant chose the target;
* where the target appeared in their ranking;
* their target rating;
* the mean control rating;
* the chance probability of a correct first choice.

For a formal study, delay the reveal until data collection for the cohort is complete. Immediate feedback can influence participants who discuss the experiment with

others.

□

8. Data recorded

Participant record

* anonymous participant ID;
* consent version;
* account ID or study token;
* age band;
* country of residence;
* astrology experience level;
* birth-data provenance;
* birth-time uncertainty;
* study cohort;
* completion status.

Birth record

* birth date;
* birth time;
* latitude;
* longitude;
* historical timezone;
* uncertainty category;
* source category.

Birth data must be stored separately from public research results.

Trial record

For each of the eleven interpretations:

* anonymous interpretation ID;
* target or control status;
* source graph ID;
* order shown;
* rating values;
* reading time;
* whether the card was revisited;

* interpretation hash;
* graph hash.

Final response

* first choice;
* second choice;
* third choice;
* confidence;
* target rank;
* target rating;
* mean control rating;
* completion duration;
* comments, where permitted.

The target/control flag should not be exposed to the participant-facing application before submission.

□

9. Primary outcome

The simplest primary outcome is binary:

* 1 if the participant selects the target interpretation;
* 0 if they select a control.

Under random choice:

$$
P(\text{correct}) = \frac{1}{11} \approx 0.0909
$$

The observed proportion of correct selections is compared with 9.09%.

Example

If 1,000 valid participants complete the experiment:

* chance predicts approximately 91 correct selections;
* the observed number is compared with that expectation;
* uncertainty intervals and an effect size are reported;
* the preregistered analysis determines whether the difference is meaningful.

A statistically significant result should not be described as proof of astrology. It would show that participants identified their target interpretations above the defined chance level under this particular procedure.

□

10. Secondary outcomes

Target rank

Record where the target appears in the participant’s full ranking.

Chance mean rank among eleven descriptions is 6.

Top-three identification

Measure how often the target appears among the top three choices.

Random chance is:

[
\frac{3}{11} \approx 27.27%
]

Rating advantage

For each participant:

[
D_i = R_{target,i} - \overline{R}_{control,i}
]

where:

* (R_{target,i}) is the rating given to the genuine interpretation;
* (\overline{R}_{control,i}) is the mean rating of the ten controls.

Confidence calibration

Compare confidence with correctness.

A real discriminatory signal should generally produce higher accuracy among more confident participants. High confidence without increased accuracy may indicate subjective conviction rather than identification.

Time-precision effect

Compare results by birth-time provenance and uncertainty.

A structural astrological hypothesis predicts that accurate timing should matter more where angles, houses and rapidly changing relations are included.

Astrological experience

Compare:

* participants with no astrological knowledge;
* casual users;
* practising astrologers.

This must be treated cautiously and preferably preregistered.

□

11. Preventing accidental clues

The application must ensure that participants cannot identify the target for non-astrological reasons.

Required controls

* Equal section count.
* Similar word count.
* Identical typography.
* No planet names.
* No zodiac signs.
* No dates.
* No locations.
* No generational references.
* No mention of rare chart configurations by name.
* No target-specific introductory language.

* No difference in generation model or prompt.
* No ordering bias.
* No target always appearing in the same position.
* No longer generation time visible for the target.
* No browser response containing hidden target metadata.
* No target flag in client-side JavaScript before submission.

The correct answer must remain server-side.

□

12. Security and integrity

Server-side generation

Birth data, target assignment and control generation occur on the server.

The browser receives only:

* anonymous interpretation IDs;
* rendered descriptions;
* rating schema;
* shuffled display order.

Submission integrity

Each trial receives:

* a unique token;
* a cryptographic hash;
* a creation timestamp;
* a submission timestamp.

After submission, the record becomes immutable.

Duplicate participation

The system should detect probable duplicate participation using privacy-conscious methods such as:

* verified email;
* one-time invitation tokens;
* study account;
* rate limits;
* duplicate birth-record warnings;
* optional device-session checks.

Device fingerprinting should not be used without explicit disclosure.

Bots

Use:

* email verification;
* minimum reading-time checks;
* attention checks;
* rate limiting;
* server-side validation;
* suspicious-response flagging.

Participants should not be excluded solely for reading quickly. Exclusion rules must be defined before analysis.

□

13. Suggested database structure

users

* id
* email_hash
* created_at
* consented_at
* consent_version
* account_status

participants

* id
* user_id
* study_id
* anonymous_code
* age_band

* country
* astrology_experience
* status

birth_records

* id
* participant_id
* date_of_birth
* time_of_birth
* latitude
* longitude
* timezone
* time_source
* time_uncertainty
* encrypted_payload

graphs

* id
* graph_hash
* engine_version

* input_record_id
* graph_json
* relation_count
* provenance_type

interpretations

* id
* graph_id
* interpretation_hash
* model_version
* prompt_version
* lexicon_version
* validator_version
* content
* validation_status
* generation_metadata

trials

* id
* participant_id
* study_version
* target_interpretation_id
* created_at
* submitted_at
* status

trial_items

* id
* trial_id
* interpretation_id
* display_position
* is_target
* overall_rating
* specificity_rating
* reading_duration

trial_results

* id
* trial_id
* first_choice_id
* second_choice_id
* third_choice_id
* confidence
* target_rank
* target_selected
* submitted_at

The is_target field must never be sent to the client before final submission.

□

14. Administration interface

The research dashboard should show:

* number of invited participants;
* number who consented;
* number who started;

* number who completed;
* valid and excluded submissions;
* model and engine versions;
* interpretation-generation failures;
* validation rejection rate;
* distribution of target positions;
* completion times;
* birth-time quality distribution;
* frozen preregistration document;
* export controls.

Before the study closes, the dashboard should avoid showing the developing success rate to anyone involved in changing the software or analysis. This prevents unconscious mid-study adjustment.

A separate sealed-results mode should reveal the primary outcome only after:

* recruitment has closed;
* exclusions have been applied according to the registered rules;

* the analysis script has been frozen;
* the dataset hash has been recorded.

□

15. Experimental variants

Variant A: Plain-language descriptions

All astrological terminology is removed.

This is the recommended primary public experiment.

Variant B: Astrological descriptions

Planet, sign and aspect names remain visible.

Suitable for testing astrologers, but participants may identify generational or familiar chart details.

Variant C: No language model

Interpretations are assembled deterministically from fixed symbolic templates.

This removes model variation and provides a valuable baseline.

Variant D: Generic-text control

Include one or more professionally written Forer-style descriptions.

This tests whether generated target readings outperform deliberately generic prose.

Variant E: Own chart versus other real charts

All controls come from real participants.

This is the strongest recommended confirmatory design.

Variant F: Birth-time sensitivity

Generate several interpretations from the same date and location but different plausible birth times.

This tests whether time-sensitive geometry contributes anything beyond slower planetary placements.

## 16. Minimum viable prototype

The first working prototype does not need the complete research infrastructure.

MVP features

* participant account or single-use token;
* consent page;
* birth-data form;
* location geocoding;
* Chronometer graph generation;
* ten matched control moments;
* eleven fixed-format interpretations;
* randomised order;
* one-to-five rating for every interpretation;
* first-choice selection;
* confidence question;
* locked submission;
* simple personal result;

* anonymised CSV export;
* version logging.

MVP research limitations

The prototype should be labelled exploratory because:

* controls may be synthetically generated;
* the model may be remotely hosted and mutable;
* sample recruitment may be self-selecting;
* participants may discuss results publicly;
* immediate result disclosure may affect later recruits;
* no formal power analysis may yet exist.

The MVP’s purpose is to test:

* whether people complete the task;
* whether eleven descriptions create too much fatigue;
* whether the descriptions are distinguishable in unintended ways;
* whether the interface works on mobile;
* whether rating and ranking data are collected reliably.

It should not be marketed as the decisive test.

□

17. Recommended build sequence

Phase 1: Internal pilot

Use 10–20 invited testers.

Test:

* generation consistency;
* mobile reading burden;
* accidental clues;
* database integrity;
* target concealment;
* result calculation.

Phase 2: Open exploratory study

Recruit approximately 100–300 participants.

Publish:

* experimental design;
* known limitations;
* anonymous aggregate results;
* all software versions.

Use the results to estimate completion rate, variance and plausible effect size.

Phase 3: Preregistered confirmatory study

Before recruitment:

* freeze the protocol;
* conduct a power analysis;
* publish the smallest effect size of interest;
* lock the engine and interpretation pipeline;
* specify exclusion rules;
* define the primary endpoint;
* define the stopping rule;
* register the analysis script.

Use matched real-participant charts wherever possible.

Phase 4: Independent replication

Give the frozen protocol and code to an independent team, ideally including sceptical researchers.

The strongest result would not be one produced by the Chronometer’s creator. It would be one reproduced by people who expected it to fail.

□

18. Success and failure conditions

Product success

The application succeeds as a product when:

* participants understand the task;
* the experience works smoothly on mobile;
* most participants complete it;
* target provenance remains concealed;
* the data can be exported and audited;
* every result can be traced to frozen graph and

interpretation versions.

Empirical success

The experiment produces evidence of an effect only if the target interpretation performs above the preregistered chance and effect-size thresholds.

Empirical failure

The specified implementation fails if:

* target selection is consistent with chance;
* target ratings do not exceed control ratings by a meaningful amount;
* or any apparent advantage disappears under replication or proper controls.

A null result must remain publishable.

The system must not respond to failure by silently changing:

* aspect definitions;
* orb thresholds;
* symbolic meanings;
* participant exclusions;
* outcome measures;
* or statistical methods.

Such changes may motivate a new experiment, but they cannot rescue the completed one.

□

19. Participant-facing summary

This experiment generates eleven anonymous descriptions from eleven different birth geometries. One uses the birth information you supplied. The other ten use matched, astronomically possible birth records.

The system receives no information about your personality, history, occupation or interests.

Your task is to rate the descriptions and choose the one that resembles you most.

If people cannot identify their own description more often than chance, the experiment provides evidence against the idea that this method extracts person-specific information from birth geometry.

If people identify their own description more often than chance, the result would justify further controlled testing. It would not, by itself, establish why the association exists.

□

20. The experiment in one sentence

Give each person one interpretation produced from their own birth geometry and ten produced from matched alternative geometries, hide which is which, and measure whether they can find themselves more often than chance.

That is the whole machine.

Eleven skies enter. The participant chooses one. The aggregate result decides whether anything is there.

*Toby Godden*