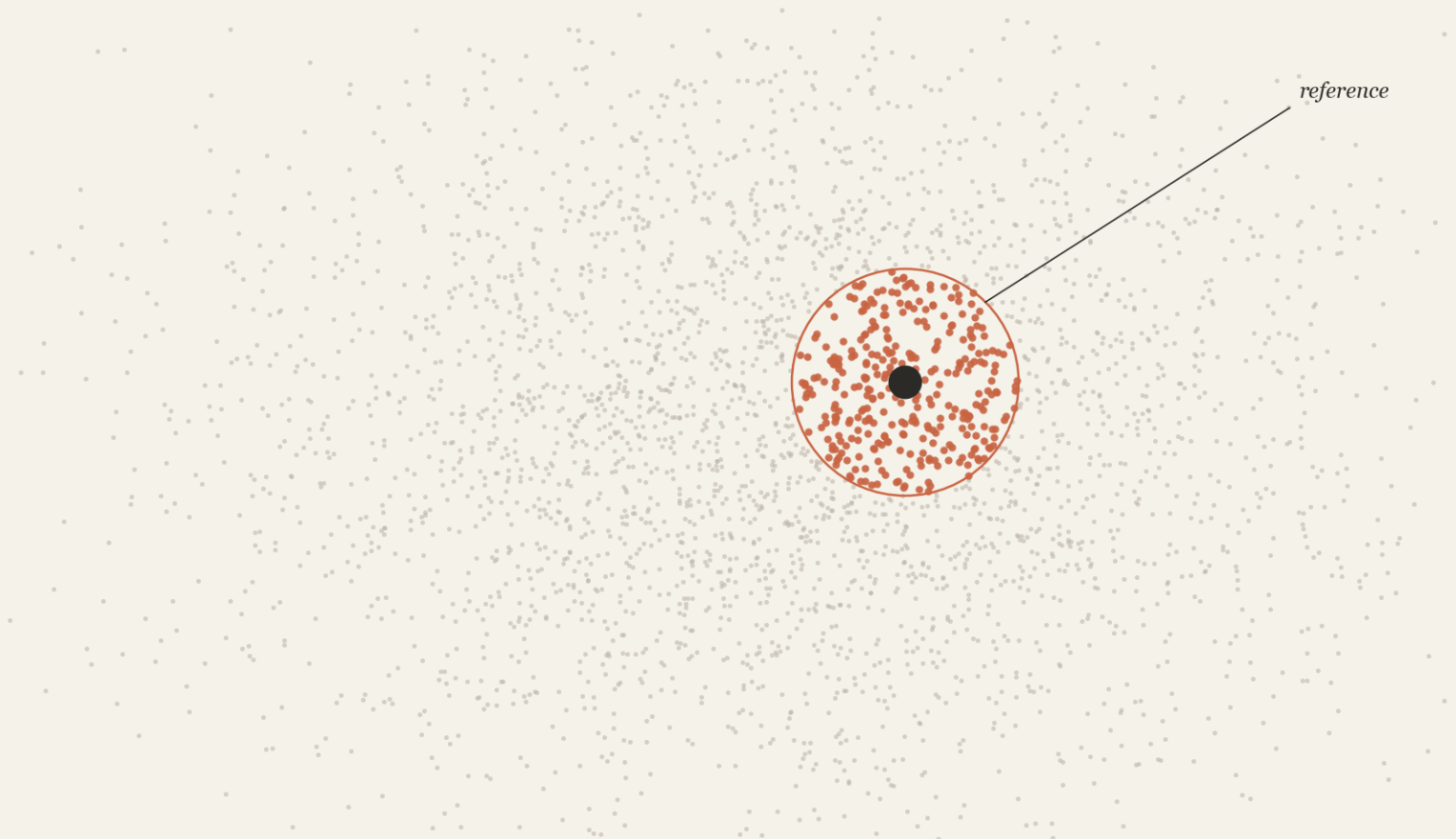


The Hunter Variant Hypothesis

Why similar faces recur: genetics, biological constraint, probability and perception in the brunette Northern and Northwestern European archetype



reference

Hunter Hughes

Founder, H Heuristics

hheuristics.substack.com

H HEURISTICS · RESEARCH REPORT

The Hunter Variant Hypothesis

Why similar faces recur: genetics, biological constraint, probability and perception in the brunette Northern and Northwestern European archetype

AUTHOR

Hunter Hughes, Founder, H
Heuristics

PUBLISHED

October 2026

FORMAT

Research report · open access

PREPRINT

[doi:10.6084/m9.figshare.34856661](https://doi.org/10.6084/m9.figshare.34856661)

ORIGINAL ESSAY

[H Heuristics on Substack](#)

SCOPE

Exploratory hypothesis and research
design

ABSTRACT

Unrelated people frequently appear remarkably similar. The Hunter Variant Hypothesis, proposed by H Heuristics, studies this recurrence from a fixed vantage point, using Hunter Hughes as a reference case for the brunette Northern and Northwestern European archetype: medium-to-dark brown hair, fair skin, light-to-intermediate eyes and an elongated-to-angular facial architecture. This report synthesizes the H Heuristics essays and preprint with peer-reviewed research in population genetics, facial genomics, developmental biology, probability, cognitive neuroscience and biometrics. It argues that recurrence emerges from four interacting layers. Genetics supplies a shared pool of common, polygenic appearance variants. Biological constraint bounds and clusters the space of faces. Statistical probability guarantees partial recurrence in a population of eight billion. And human perception, which is holistic, cue-driven and threshold-based, converts potential resemblance into experienced resemblance, amplified by online exposure and the Pareto distribution of visibility. The report formalizes the five-type Hunter Variant Spectrum, quantifies the archetype's likely prevalence with illustrative models, reviews evidence that algorithmically confirmed look-alikes share genetic variation, and specifies a pre-registered, blind, multi-rater research protocol with matched controls, algorithmic cross-checks, explicit null-result interpretations, ethical safeguards and an eighteen-month roadmap.

SUGGESTED CITATION

Hughes, H. (2026). *The Hunter Variant Hypothesis: Why Similar Faces Recur — Genetics, Biological Constraint, Probability and Perception in the Brunette Northern and Northwestern European Archetype*. H Heuristics Research Report.

A NOTE ON THE FIGURES

Figures marked *illustrative*, *schematic* or *simulated* present models and design examples, not measured results. Reference images are reproduced with the subject's consent. Third-party images in the author's archive remain the property of their rights holders, and the individuals shown are not identified. © 2026 Hunter Hughes / H Heuristics.

IN THIS REPORT

Contents

	Executive Summary	5
	The argument in brief, key findings and the contribution of this report	
01	The Observation and the Question	7
	From a familiar experience to a structured question; the three founding propositions	
02	The Reference Case: Defining the Hunter Archetype	10
	Defining the reference appearance; biological core versus presentational layer	
03	The Variant Spectrum and the Hybrid Hunter	14
	Partial resemblance and the five-type Hunter Variant Spectrum	
04	The Brunette Northern and Northwestern European Archetype	18
	Population history, deep ancestry and why the archetype is common	
05	The Genetics of Appearance I: Pigmentation	21
	Polygenic coloring, HERC2/OCA2, MC1R and the brunette-with-light-eyes combination	
06	The Genetics of Appearance II: Facial Shape	24
	Heritability, facial GWAS and convergent genetic recipes	
07	Biological Constraints: Why Faces Converge	27
	Developmental programs, integration, modularity and averageness	
08	Statistical Probability: The Mathematics of Recurrence	29
	Combinatorics, the archetype funnel and the birthday problem for faces	
09	Human Perception: How Resemblance Is Recognized	33
	Holistic processing, cue salience, face space and the variant zone	
10	Machines and Look-Alikes: Biometrics and Genetic Evidence	37
	Face embeddings, false matches and the 2022 look-alike genetics study	
11	The Internet as Variant Amplifier and the Pareto Distribution of Visibility	40
	Exposure, recommendation feedback loops and alpha variants	
12	The Variant Matrix and the "Variant Gene" Metaphor	44
	The wider landscape of archetypes and the metaphorical variant gene	
13	Testing the Hypothesis: A Research Design	47
	Testable claims, a six-step protocol and what a null result would mean	
14	Ethics, Privacy and the Limits of the Framework	51
	Privacy, consent, image rights and safeguards against typological misuse	
15	Implications for H Heuristics and a Research Roadmap	54
	Strategic fit, communication and an eighteen-month research plan	
	Conclusion: A Recurring Human Form	57
	Four layers, two conclusions and the open questions	
A	Glossary of Terms	58
	Definitions of key terms	
B	Resemblance Scoring Instrument	59
	Rater form and pre-specified classification rules	
	References	60
	Sources cited, with links	

VISUAL INDEX

Figures and Tables

Figure 1	The original H Heuristics framing of the Hunter Variant Hypothesis	7
Figure 2	Reference images of the subject, Hunter Hughes	10
Figure 3	Stability of appearance dimensions: biological core versus presentational layer	12
Figure 4	The Hunter Variant Spectrum: five provisional types of resemblance	14
Figure 5	Illustrative resemblance profiles for three variant types across the five dimensions	16
Figure 6	The author's informal "Variants" album (January–November 2025, 15 items)	17
Figure 7	Approximate ancestry composition across European regions (schematic)	19
Figure 8	Hair color as a polygenic continuum: schematic distributions for two European populations	21
Figure 9	Growth in the number of genomic loci linked to normal facial shape variation	25
Figure 10	Expected number of matching individuals among 8.1 billion people, by number of required traits	29
Figure 11	Illustrative funnel: how many people might share the defining traits of the Hunter archetype	30
Figure 12	Probability of encountering at least one perceived look-alike of a given reference, by exposure	31
Figure 13	Face-space illustration: the number of near neighbors depends on where the reference sits	34
Figure 14	Signal detection view of resemblance: unrelated pairs that cross the recognition threshold form the "variant zone"	35
Figure 15	The variant amplifier: engagement-driven recommendation can inflate perceived prevalence	41
Figure 16	The Pareto distribution of visibility: a small share of people receive most of the attention	42
Figure 17	A simplified two-dimensional slice of the variant matrix	44
Figure 18	Six-step protocol for testing the Hunter Variant Hypothesis	48
Figure 19	Illustrative analysis using simulated ratings for three comparison groups	49
Figure 20	Proposed research roadmap for the Hunter Variant program	54
Table 1	The four layers of the Hunter Variant Hypothesis at a glance	6
Table 2	The three founding propositions of the Hunter Variant Hypothesis	8
Table 3	Structure of the report by analytical layer	9
Table 4	Operational coding scheme for the Hunter reference archetype	13
Table 5	Major pigmentation loci relevant to the brunette Northern European archetype	22
Table 6	Metaphorical versus literal readings of key H Heuristics terms	45
Table 7	Core claims of the Hunter Variant Hypothesis and how each could be tested	47
Table 8	Ethical safeguards for the Hunter Variant research program	52
Table 9	Glossary of key terms used in this report	58
Table 10	Hunter Variant Resemblance Scoring Instrument (rater form)	59
Table 11	Pre-specified classification rules (illustrative thresholds to be fixed at pre-registration)	59

OVERVIEW

Executive Summary

The Hunter Variant Hypothesis begins with an observation most people have had: two people with no known connection can look remarkably alike. A stranger on a train, an actor in a period drama, a soldier in a colorized archive photograph or a musician on a streaming platform can suddenly call to mind someone we know. H Heuristics formalizes that experience by fixing a single reference point. It uses its founder, Hunter Hughes, as a documented reference case and asks how a recognizable appearance, the brunette Northern and Northwestern European archetype in which he sits, appears again and again across unrelated individuals.

This report is the first full-length treatment of the hypothesis. It brings together the original H Heuristics essays ([The Hunter Variant Hypothesis](#), [Hunter Hughes](#), [H Heuristics](#), and [the Hunter Variant Phenomenon](#) and [The Variant \(or Doppelgänger\) Phenomenon](#)) and the deposited preprint ([figshare](#), [doi:10.6084/m9.figshare.34856661](https://doi.org/10.6084/m9.figshare.34856661)). It sets them against the peer-reviewed literature in population genetics, facial genomics, statistics, cognitive neuroscience and biometrics. The goal is not to prove that "Hunter variants" exist as a biological category. They do not, and the hypothesis has never claimed that they do. The goal is to explain why the impression of recurrence is predictable, and to set out how it could be measured.

The argument in brief

The report argues that four layers interact to produce the perceived recurrence of a specific human appearance:

- **Genetics.** Hair color, eye color and facial shape are polygenic traits, each shaped by hundreds of genetic variants of small effect. In Northern and Northwestern Europe, the allele frequencies that produce medium-to-dark brown hair together with light or intermediate eyes and fair skin are common. The traits that define the Hunter archetype are therefore drawn from a large, shared pool rather than from a rare one.
- **Biological constraint.** Faces are built by conserved developmental programs. The space of viable human faces is vast but bounded, and most people cluster near its center. Convergent combinations are expected.
- **Statistical probability.** With roughly 8.1 billion people alive, even combinations with a probability of one in a million recur thousands of times. A rough count shows that a strict, six-filter definition of the Hunter archetype could still describe well over a hundred thousand people in Northern and Northwestern Europe alone.
- **Human perception.** People do not compare faces point by point. They process faces holistically, rely on a handful of salient cues such as hair, brows and face outline, and are surprisingly poor at matching unfamiliar faces. A partial resemblance of perhaps a quarter to a half of the cues is often enough to trigger a sense of familiarity.

Key Findings

- The Hunter Variant Hypothesis is best understood as a *resemblance* hypothesis, not an *identity* or *ancestry* hypothesis. Look-alikes need share no special connection beyond ordinary human variation.
- The brunette Northern and Northwestern European archetype is common because its component traits are common. Recurrence is the expected result of polygenic inheritance in populations with shared deep ancestry.
- A [2022 study in Cell Reports](#) found that unrelated look-alikes identified by facial-recognition algorithms shared more genetic variants than random pairs. This suggests that some striking resemblances partly reflect a shared genetic "recipe" rather than chance alone.
- The internet acts as a *variant amplifier*. It expands the number of faces any observer sees by orders of magnitude, and recommendation systems feed back on what people engage with. Perceived frequency is not measured prevalence.
- The hypothesis is testable. This report specifies a pre-registered, blind, multi-rater protocol with matched brunette controls and algorithmic cross-checks. It also defines what a null result would mean.

What this report contributes

Earlier H Heuristics writing introduced the five-type *Hunter Variant Spectrum*, the concept of the *hybrid Hunter*, the *variant matrix* and the idea of *alpha variants*. These are visible people whose appearance is overrepresented because attention follows a [Pareto-like distribution](#). This report adds four things. It places the archetype in its population-genetic context. It makes the probability argument quantitative. It grounds the perceptual argument in face-recognition science. And it turns the hypothesis into a concrete research program with an eighteen-month roadmap, a scoring instrument and explicit ethical safeguards.

Throughout, the report keeps two registers apart. One is the established science of human appearance and perception. The other is the H Heuristics interpretive framework built on that science. Where H Heuristics proposes a term, a category or a rough heuristic, such as the "25 to 50 percent likeness" shorthand, the report labels it as such. Where it uses an illustration or a simulation, the figure is marked accordingly. That discipline is what separates a hypothesis that invites evidence from a claim that resists it.

TABLE 1 The four layers of the Hunter Variant Hypothesis at a glance

Layer	Core mechanism	Strongest evidence	Implication for variants
Genetics	Polygenic inheritance of pigmentation and facial shape	GWAS of hair color and face shape; 2022 look-alike genetics study	Component traits are common and shared across the region
Biological constraint	Conserved developmental programs; integration and modularity	Developmental biology; averageness research	Faces cluster; correlated features create convergent resemblance
Statistical probability	Combinatorics across 8.1 billion people	Birthday-problem logic; metric uniqueness studies	Partial matches are inevitable; full matches are rare
Human perception	Holistic, cue-driven recognition with a variable threshold	Face-space theory; familiar vs. unfamiliar face matching	Observers convert partial overlap into felt familiarity

01

The Observation and the Question

A familiar experience, an unfamiliar framing

Almost everyone has been told, at some point, that they look like someone else. The experience is so common that it has a long cultural history. The *doppelgänger* of German folklore was an uncanny double. Nineteenth-century novels used mistaken identity as a plot device. Modern social media has whole genres of "twin stranger" content. What is unusual about the Hunter Variant Hypothesis is not the observation but the framing. Instead of treating look-alikes as isolated curiosities, it treats them as data points in a pattern and anchors that pattern to a single, documented reference person.

The original essay frames the question carefully. It asks whether these similarities are isolated coincidences, predictable outcomes of shared human genetics, or part of broader statistical patterns in how human appearances are distributed. Those three options are not mutually exclusive. A central claim of this report is that all three operate at once. Some resemblances are coincidental. Many are the predictable result of shared genetic architecture. And the *visibility* of resemblance follows statistical laws of attention that have little to do with biology.

FIGURE 1 *The original H Heuristics framing of the Hunter Variant Hypothesis*

The Hunter Variant Hypothesis

An H Heuristics thesis by Hunter Hughes

The Hunter Variant Hypothesis can be developed as a framework for examining a recurring human phenomenon: the apparent reappearance of similar faces, physical features, hairstyles, and overall appearances across different individuals. It asks whether these similarities are merely isolated coincidences, predictable outcomes of shared human genetics, or part of broader statistical patterns in how human appearances are distributed.

The hypothesis fits naturally within H Heuristics' wider interest in recurring structures, mathematical regularities, and the underlying architecture of reality. — H Heuristics

Its central premise is that **human individuality may coexist with recurring patterns of physical form**. Although every person has a distinctive biological identity and life history, the combinations of traits that produce a recognizable human appearance may recur across unrelated people more often than intuition would suggest.

The hypothesis treats these recurrences as a subject for systematic investigation rather than assuming that visually similar people share a hidden connection.

Source: H Heuristics, "The Hunter Variant Hypothesis," hheuristics.substack.com (2026)

The central premise: individuality coexists with recurring form

The hypothesis rests on a deceptively simple premise. Human individuality may coexist with recurring patterns of physical form. Every person has a distinct genome (identical twins aside), a distinct developmental history and a distinct life story. Yet the *visible* output of all that individuality, the face we present to the world, is drawn from a limited vocabulary of traits: hair color and texture, eye color, skin

tone, the proportions of the face, the shape of the brow, nose, lips and jaw. When these traits combine in similar ways, the result can be a face that reads as familiar even though the person behind it is entirely unrelated.

This premise has an important consequence. The hypothesis does not need, and does not propose, any hidden mechanism linking look-alikes. As the [core thesis](#) puts it, the important distinction is between resemblance and identity. Two people can look remarkably alike without being genetically identical, closely related or connected through any identifiable mechanism beyond ordinary human variation.

Three founding propositions

The H Heuristics thesis builds on three related propositions. Each maps onto a body of established science that later chapters examine in depth.

TABLE 2 *The three founding propositions of the Hunter Variant Hypothesis*

Proposition	H Heuristics formulation	Scientific grounding (chapter)
Recurrence	Similar human appearances emerge independently across populations.	Population genetics; probability (Ch. 04, 08)
Convergence	Different combinations of inherited and developmental traits can produce strikingly similar overall appearances.	Facial genomics; developmental constraint (Ch. 06, 07)
Recognizability	Certain combinations of facial structure, hair, proportions and other characteristics are especially likely to trigger recognition of a particular person in another.	Face perception; cue salience (Ch. 09, 10)

The *recurrence* proposition is about frequency. It claims that similar appearances arise again and again without any common cause beyond the shared gene pool. The *convergence* proposition is subtler. It recognizes that two faces can look alike for *different* underlying reasons. One person's long face might come from one set of genetic variants and another's from a completely different set, much as two different arithmetic paths can reach the same number. The *recognizability* proposition shifts attention from the faces to the observer. Some trait combinations are simply more likely to set off the brain's sense of familiarity.

Why a reference person?

Studies of facial resemblance usually work with pairs: twin A and twin B, or stranger A and stranger B. The Hunter Variant Hypothesis instead fixes one point and studies the cloud of faces around it. This has real methodological advantages. A fixed reference makes the comparison repeatable. The same reference images can be shown to many raters, compared against many candidates and encoded by many algorithms. It also allows the gradual, partial resemblances that pair-based studies ignore to be mapped as a spectrum rather than forced into a yes-or-no verdict.

Using Hunter Hughes as that reference has a second rationale specific to H Heuristics. The organization is named after its founder, and its wider program examines "recurring structures, mathematical regularities and the underlying architecture of reality," in the words of the original essay. A recurring human form is a natural object of study for an institution with that remit, and the founder is the most accessible, consenting and well-documented reference subject available. As the companion essay notes, the purpose

is not to claim ownership over other people or their appearances. It is to study a pattern from a stable vantage point.

Scope of this report

The broader H Heuristics framework, the *Human Variant* or *variant matrix*, is explicitly population-general. The [Variant \(or Doppelgänger\) Phenomenon](#) essay stresses that every broad human population is likely to contain its own recurring archetypes, and that the idea should not be restricted to Europeans or to any one ancestry. This report narrows the lens deliberately. It focuses on the archetype in which the reference case sits: the *brunette Northern and Northwestern European* appearance, characterized by medium-to-dark brown hair, fair skin, light or intermediate eye color and the facial proportions described in Chapter 02.

That focus is analytical, not hierarchical. It reflects where the reference case is located in the space of human appearance, and it lets the report use the unusually rich genetic literature on European pigmentation and facial shape. Nothing in the analysis implies that this archetype is more significant, more desirable or more "original" than any other. The same tools would apply to any archetype in any population.

How this report is organized

The chapters follow the logic of the hypothesis itself. Chapters 02 and 03 define the reference case and the spectrum of resemblance. Chapters 04 to 07 explain where the component traits come from and why faces converge. Chapters 08 to 11 explain how often recurrence should occur and why it is noticed, amplified and concentrated among visible people. Chapters 12 to 15 situate the hypothesis in the wider H Heuristics framework, turn it into a testable research design, set its ethical boundaries and outline a research roadmap.

TABLE 3 *Structure of the report by analytical layer*

Part	Chapters	Question addressed
Definition	02 Reference case · 03 Variant spectrum	What exactly is being recognized, and at what depth?
Origins	04 Population context · 05 Pigmentation · 06 Facial shape · 07 Constraint	Where do the archetype's traits come from, and why do faces converge?
Recurrence and recognition	08 Probability · 09 Perception · 10 Machines · 11 Amplification	How often should variants occur, and why are they noticed?
Framework and method	12 Variant matrix · 13 Testing · 14 Ethics · 15 Roadmap	How can the hypothesis be tested responsibly, and what follows?

In brief

The Hunter Variant Hypothesis treats look-alikes as a pattern to be measured rather than a coincidence to be marveled at. Its three propositions, recurrence, convergence and recognizability, each correspond to established science. Its use of a fixed reference person turns an anecdote into a repeatable comparison.

02

The Reference Case: Defining the Hunter Archetype

Why the reference must be specified

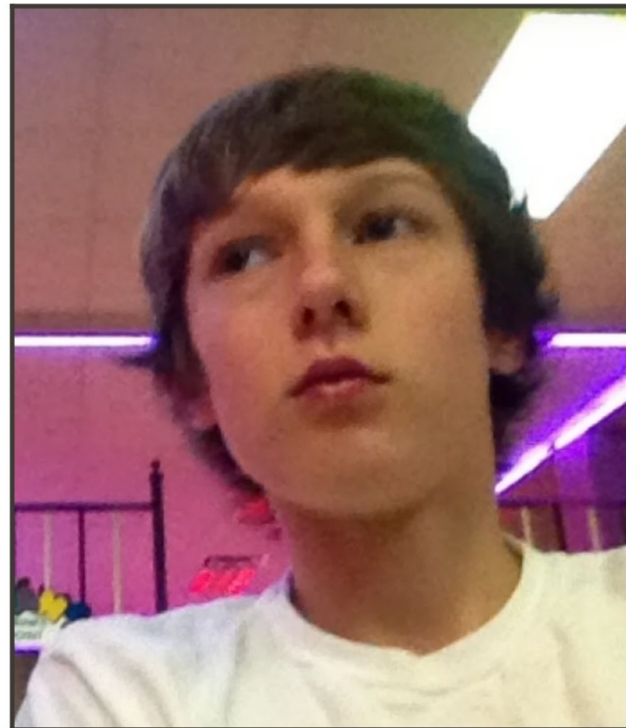
A hypothesis about resemblance to a person is only as good as its description of that person. If the reference is vague, almost anyone can be called a match. If it is overly precise, no one can. The H Heuristics essays therefore define the reference case along explicit dimensions and invite each of them to be scored separately. This chapter consolidates that definition and documents it with reference images supplied by the subject.

FIGURE 2 *Reference images of the subject, Hunter Hughes*

A Adult reference image



B Adolescent reference image



Source: Supplied by the subject for this report. Images are reproduced with consent and are not to be reused without permission.

The defining features

The [original essay](#) describes the reference appearance in five clusters. Drawing on that description and the reference images above, the Hunter archetype can be specified as follows.

- **Hair and coloring.** Medium-to-dark brown hair with a slight natural wave, worn long on top and swept back with a prominent, separately falling front section. The complexion is fair, and in the adult reference image the eyes read as light-to-intermediate in tone. Facial hair, when present, forms a light-to-medium beard concentrated on the chin and upper lip.
- **Facial structure.** An elongated-to-angular face with a high forehead, a relatively long mid-face and a pronounced, defined jawline. The proportions are vertically oriented rather than broad.
- **Feature configuration.** Straight, defined eyebrows set fairly low over the eyes, a straight and moderately long nose, and medium-full lips. The spacing reads as balanced rather than wide-set or close-set.
- **Presentation.** Grooming and styling choices, above all the swept-back hairstyle, which the series has treated as the subject's signature, along with typical clothing, posture and expression.
- **Composite resemblance.** The overall gestalt, or the impression that a person "reads as" the reference even when individual features differ.

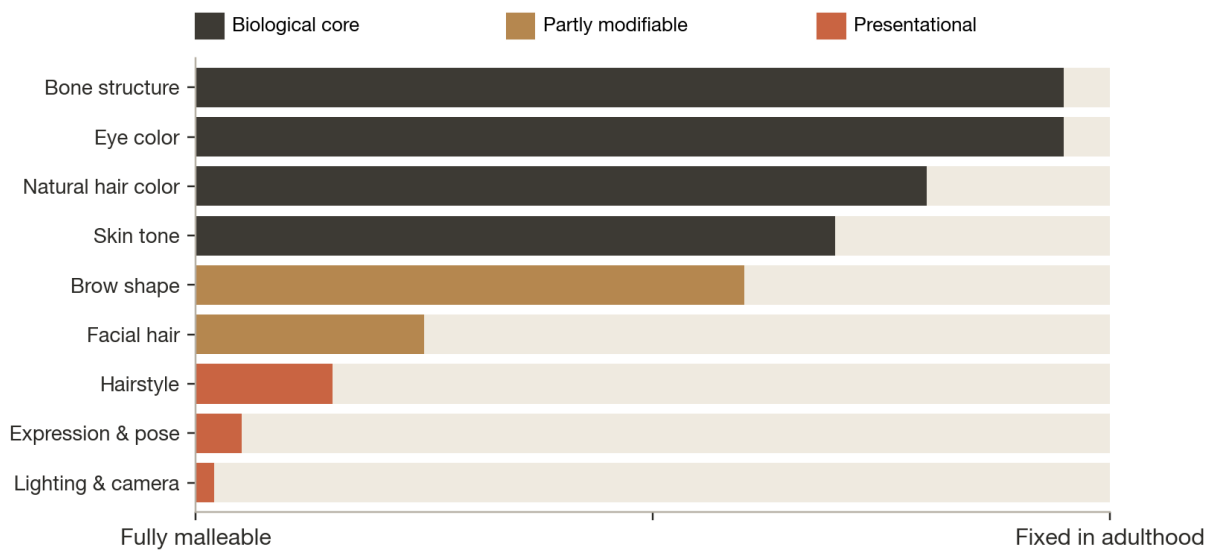
Developmental continuity

The adolescent reference image adds a dimension that is often overlooked. Resemblance is age-dependent. In adolescence, the same individual presents a rounder face, fuller cheeks, a heavier fringe and less defined brow and jaw structure. The coloring is continuous: brown hair, fair skin and the same nasal profile. The skeletal proportions that define the adult archetype, however, have not yet fully emerged.

This matters for the hypothesis in two ways. First, it means that the "Hunter archetype" is really an *adult* archetype. A candidate variant should be compared with a reference of similar age, or the comparison will be confounded by developmental differences. Second, it shows that one person can sit in different regions of appearance space at different life stages. The [Variant Phenomenon essay](#) makes the same point when it notes that age, weight, facial hair and expression can make one person resemble different variants at different times.

Biological core versus presentational layer

Not all of the defining features are equally stable. Some, such as bone structure and eye color, are effectively fixed in adulthood. Others, such as natural hair color and skin tone, change slowly with age and sun exposure. Still others, such as hairstyle, facial hair, clothing, expression, lighting and camera angle, can change in minutes. A rigorous hypothesis has to keep these layers apart. A resemblance built mostly on presentation is a different phenomenon from one built on facial architecture, even though observers may report both as "looking like Hunter."

FIGURE 3 *Stability of appearance dimensions: biological core versus presentational layer*

Qualitative stability index for illustration

Source: H Heuristics analysis. Qualitative index for illustration.

This distinction carries real weight. It is the basis of the Type IV, or *Stylistic*, category in the variant spectrum (Chapter 03). It also explains a common observation: resemblance can appear and disappear as a person changes hairstyle or grows a beard. Because the swept-back hairstyle is such a strong visual cue, a sizeable share of perceived Hunter variants may be driven partly by presentation. A testing protocol therefore needs to measure resemblance both with and without hair visible (Chapter 13).

The reference as observer

Using the founder as the reference case brings a methodological strength and a matching weakness. The strength is access. The reference subject can provide standardized images, developmental photographs and informed consent, and can take part in repeated comparisons over time. Few doppelgänger studies have a reference subject so willing to be measured.

The weakness is that the person who defines the reference is also, often, the person who notices the variants. Self-perception of one's own face is unusual. People see their own face mostly in mirrors, which reverse it left to right, and in their own photographs, which they select and remember selectively. Research on self-face recognition suggests that people hold an idealized or habitual representation of their own appearance, which may differ systematically from how others see them. A variant that strikes the reference subject as a close match may therefore strike independent raters as weaker, or the reverse. For this reason the testing protocol separates three kinds of judgment: the reference subject's own nominations, ratings from people who know him, and ratings from people who have never seen him. Agreement across all three is the strongest evidence a resemblance can have.

An operational definition

For research purposes, the reference case can be turned into a coding scheme. The scheme below takes the H Heuristics dimensions and adds explicit, observable sub-criteria that blind raters can score. It is the basis of the instrument in Appendix B.

TABLE 4 *Operational coding scheme for the Hunter reference archetype*

Dimension	Observable sub-criteria	Layer
Hair & coloring	Medium–dark brown hue; slight wave; fair complexion; light–intermediate eyes	Biological / partly modifiable
Facial structure	Elongated-to-angular outline; high forehead; long mid-face; defined jaw	Biological core
Feature configuration	Straight low brows; straight moderately long nose; medium-full lips; balanced spacing	Biological core
Presentation	Long swept-back hair with front section; light beard; typical expression	Presentationals
Composite	Holistic "reads as" judgment, rated last and independently	Perceptual

In brief

The Hunter archetype is defined by the medium-to-dark brown hair, fair coloring, elongated-to-angular face, defined brows and jaw, and a distinctive swept-back presentation. Separating the stable biological core from the malleable presentational layer is essential, because the two produce resemblances of quite different kinds.

03

The Variant Spectrum and the Hybrid Hunter

Partial resemblance as the central pattern

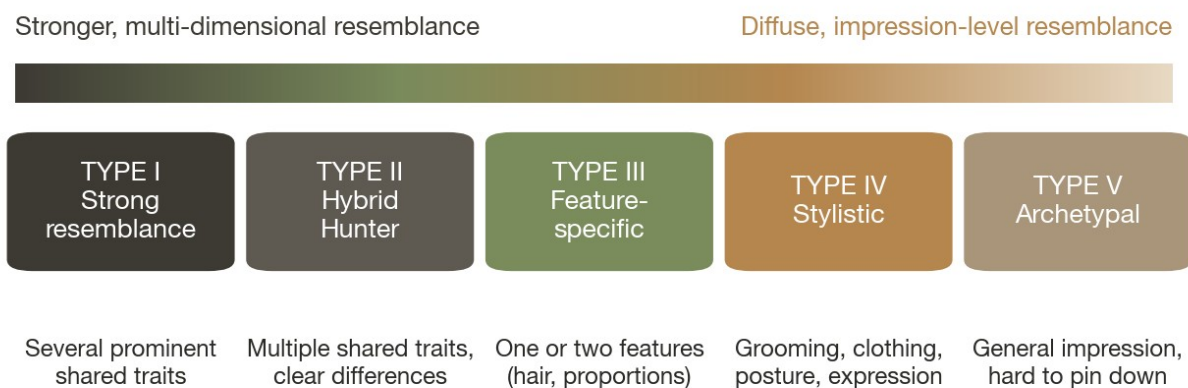
The most distinctive conceptual move in the H Heuristics framework is its emphasis on *partial* resemblance. Popular doppelgänger culture prizes the near-perfect match, the stranger who could pass as a twin. The Hunter Variant Hypothesis argues that such cases are the rare tip of a much larger distribution. Far more common is the *hybrid Hunter*: a person who shares several of the reference's traits, perhaps the hair, the brow line and the facial length, while differing clearly in others.

The [original essay](#) gives this an informal number. A hybrid variant is someone who appears to carry roughly 25 to 50 percent of the reference's image and likeness. The [companion essay](#) is explicit that this is an "initial descriptive shorthand" rather than a measured statistic. The report keeps that framing. The percentage range is a useful way to communicate the *idea* of partial overlap, and Chapter 13 explains how it can be replaced by a pre-defined, reproducible score.

The five-type spectrum

To organize partial resemblance, H Heuristics proposes a five-type *Hunter Variant Spectrum*. The types run from strong, multi-dimensional resemblance to diffuse, impression-level resemblance. The essays describe these categories as provisional, and they are best read as hypotheses about how resemblance clusters rather than as fixed taxonomic boxes.

FIGURE 4 *The Hunter Variant Spectrum: five provisional types of resemblance*



Source: After H Heuristics, "The Hunter Variant Hypothesis" (2026), Section 3.

Type I — Strong Hunter resemblance

Several prominent characteristics are shared at once, typically coloring, facial structure and feature configuration. Type I candidates are the classic "could be his brother" cases. They should be the rarest category, and they are the most likely to show the shared-genetics signal described in Chapter 10.

Type II — Hybrid Hunter

Multiple traits are shared, but there are substantial differences elsewhere. A Type II variant might share the hair, brows and facial length but have a broader nose or a rounder jaw. This is the category the H Heuristics framework treats as most characteristic of the phenomenon.

Type III — Feature-specific Hunter

The resemblance is concentrated in one or two features, most often hair color and style or facial proportions. Type III cases are common and often fleeting, and they tend to rely on a dominant cue.

Type IV — Stylistic Hunter

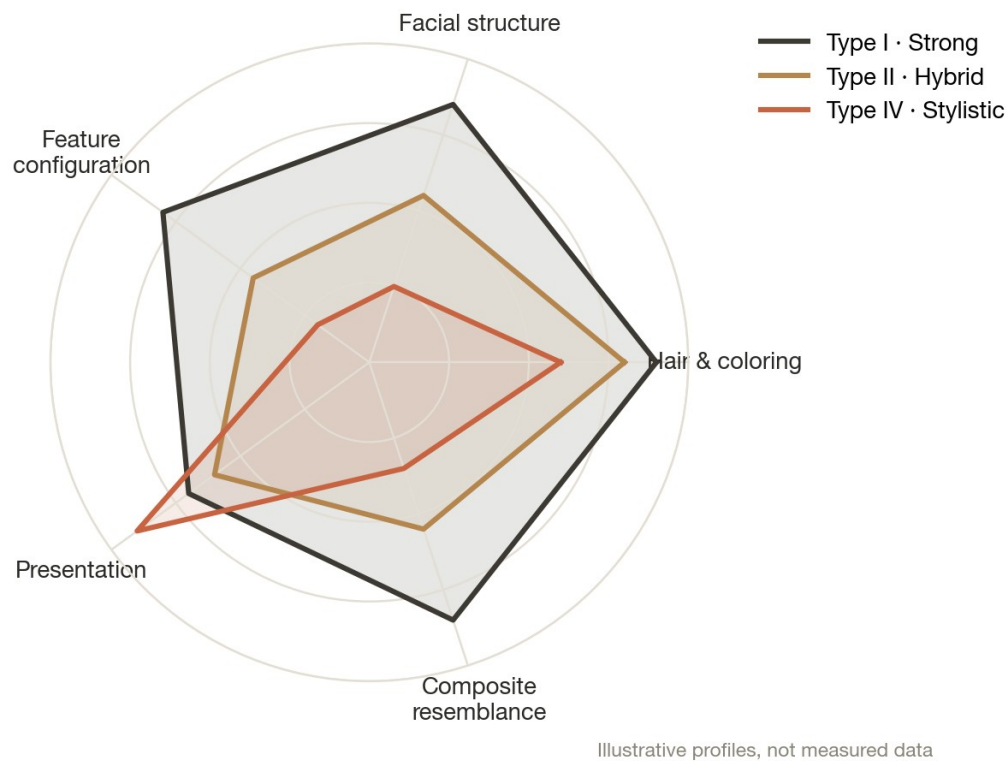
The resemblance is driven mainly by presentation: grooming, clothing, posture or expression. Because the reference's swept-back hairstyle is distinctive, Type IV cases can be striking in photographs while disappearing in person or under different styling.

Type V — Archetypal Hunter

A general impression of resemblance whose specific basis is hard to pin down consistently. Type V is the category most dependent on the observer, and it is the most vulnerable to expectation effects (Chapter 09). It is also scientifically interesting, because it may reflect the holistic, configural processing that the brain uses for faces.

Profiles across the five dimensions

One useful way to see the spectrum is to plot each type across the five analytical dimensions introduced in Chapter 02. A Type I profile is large and balanced. A Type II profile is moderately large but uneven. A Type IV profile is dominated by presentation. The figure below illustrates those shapes. The profiles are illustrative rather than measured, but they show how a multi-dimensional scoring instrument can tell types apart that a single "how similar?" rating would merge.

FIGURE 5 *Illustrative resemblance profiles for three variant types across the five dimensions*

Source: H Heuristics analysis. Illustrative profiles, not measured data.

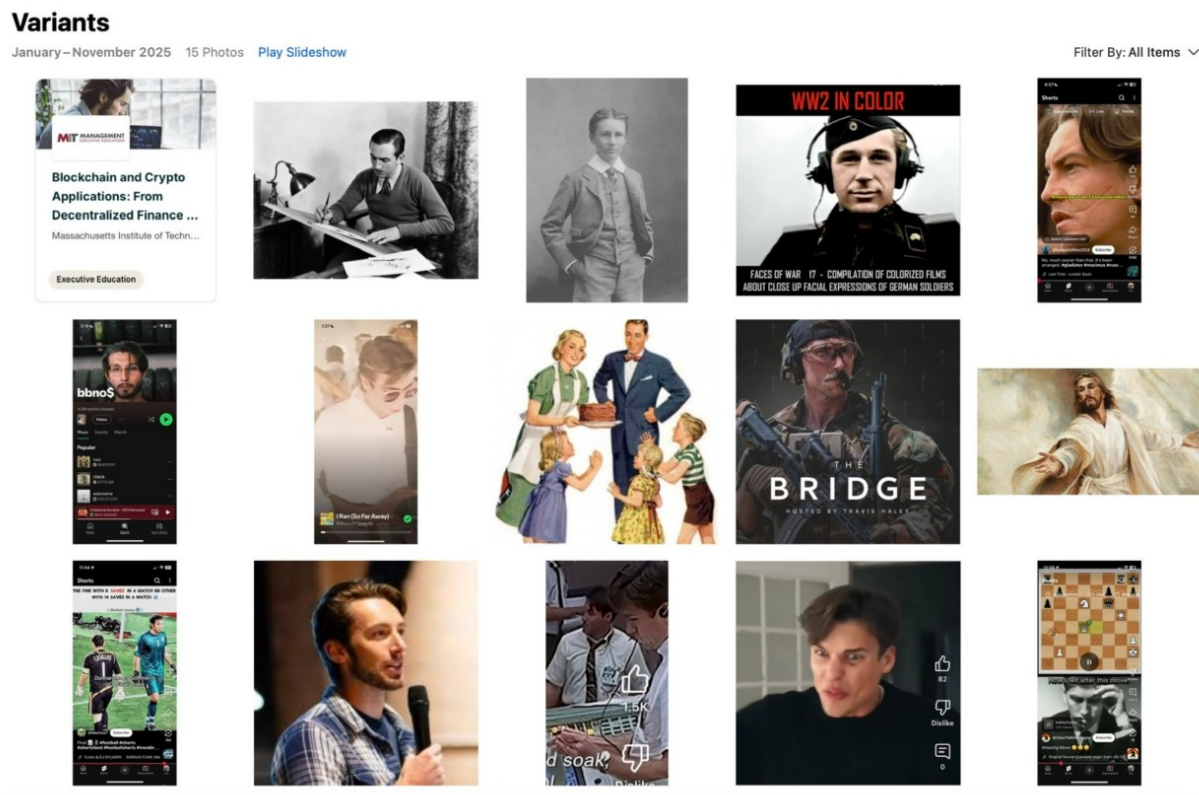
Why partial resemblance dominates

The prevalence of hybrid and feature-specific cases is not an accident. It follows from the combinatorics discussed in Chapter 08. If a full match requires agreement on many independent traits, full matches will be exponentially rarer than partial ones. Each additional trait that must match divides the pool. A partial match on three or four high-salience traits, by contrast, will be common in any population where those traits are common. In a brunette-majority population with many fair-skinned, light-eyed individuals, the coloring component of the Hunter archetype is shared by many millions of people. The rarer components are the specific facial architecture and the distinctive presentation.

There is also a perceptual reason. As Chapter 09 explains, unfamiliar faces are recognized mainly through a few external and high-contrast cues: hairline and hairstyle, face outline, and the brow-eye region. A face that matches on those cues will often *feel* like a strong match even if it differs on finer internal details. The hybrid Hunter is therefore partly a product of how human vision economizes.

From spectrum to dataset

The spectrum becomes scientifically useful once each candidate is coded consistently. The companion essay proposes a structured database recording the source of each image, the visible characteristics, the assigned category and the coder's confidence. The author's own *Variants* collection, a personal album of fifteen items gathered between January and November 2025, is an early, informal version of such an archive. It is reproduced below as a document of the hypothesis's origins, not as evidence of resemblance. The individuals shown are not identified in this report, and no claim about any of them is made.

FIGURE 6 The author's informal "Variants" album (January–November 2025, 15 items)

Source: Screenshot of the author's personal photo collection. Third-party images remain the property of their rights holders and are shown for documentary purposes only. Individuals are not identified.

The album shows exactly the sampling problem that the testing protocol is designed to solve. It mixes historical photographs, colorized archival film, religious art, illustration, sports broadcasts, podcast artwork and social-media screenshots. These images arrived through very different channels, under very different lighting and styling, and were selected *because* they seemed to resemble the reference. That is the right way to generate a hypothesis and the wrong way to test one. Chapter 13 sets out how the transition from archive to evidence can be made.

In brief

Partial resemblance, the *hybrid Hunter*, is the heart of the phenomenon. The five-type spectrum organizes resemblance by its depth and its basis, from strong multi-dimensional matches to purely stylistic or archetypal impressions. Partial matches dominate because full matches require many independent traits to coincide, while perception relies on only a few.

04

The Brunette Northern and Northwestern European Archetype

Locating the archetype

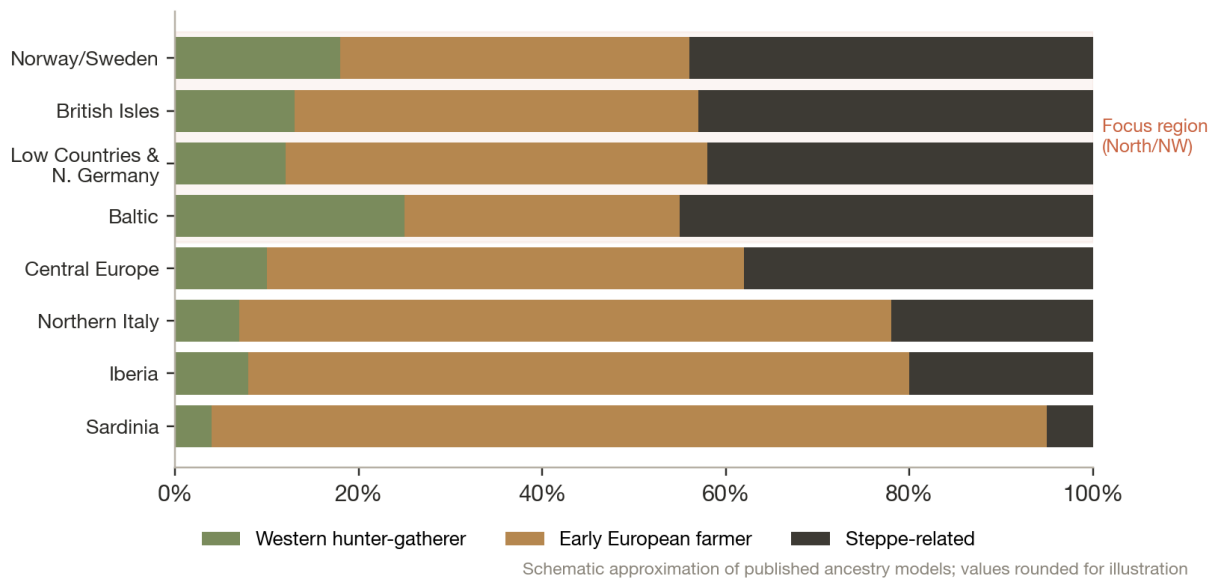
The Hunter archetype does not float free of population history. Its component traits, brown hair, fair skin and light or intermediate eyes, are found together at their highest frequencies in a particular part of the world: Northern and Northwestern Europe. That band runs from the British Isles and Ireland through the Low Countries, northern Germany and Denmark to Scandinavia and the Baltic. Understanding why these traits cluster there is the first step in understanding why the archetype recurs.

The key point is counter-intuitive. Popular imagery associates Northern Europe with blond hair. Blond hair is indeed more frequent there than almost anywhere else, but in most Northern and Northwestern European populations, *brown* shades remain the most common hair color in adulthood. In the [UK Biobank](#) cohort analyzed by [Morgan and colleagues \(2018\)](#), a large majority of white British participants reported light or dark brown hair, with blond, red and black hair as minorities. The "brunette Northern European" is therefore not a rare exception to a blond norm. It is one of the region's modal appearances.

Deep ancestry: three streams

Modern population genetics has rewritten the history of European ancestry. Ancient-DNA studies show that most present-day Europeans descend from a mixture of three ancestral streams, though in different proportions in different regions ([Lazaridis et al., 2014](#); [Haak et al., 2015](#)):

- **Western hunter-gatherers**, the post-glacial foragers of Europe. Genetic reconstructions suggest that many carried variants associated with darker skin and light eyes, an unusual combination by today's standards.
- **Early European farmers**, who spread from Anatolia from around 8,000 years ago. They carried many of the light-skin variants now common across Europe, and the large majority had brown eyes.
- **Steppe-related pastoralists**, who arrived from the Pontic–Caspian steppe around 5,000 years ago and contributed a large share of ancestry to Northern and Central Europe.

FIGURE 7 Approximate ancestry composition across European regions (schematic)

Source: H Heuristics schematic, simplified after Lazaridis et al. (2014) and Haak et al. (2015). Values are rounded approximations for illustration and vary by study and model.

Northern and Northwestern European populations carry substantial contributions from all three streams, with relatively high steppe-related and hunter-gatherer components compared with Southern Europe. Selection scans on ancient genomes ([Mathieson et al., 2015](#)) show that variants associated with lighter skin and hair rose in frequency over the last several thousand years. They point to selection on pigmentation genes that combined with this admixture to produce the distinctive pigmentation profile of the region.

Genes mirror geography

Within Europe, genetic variation is strongly structured by geography. In a landmark analysis of about 3,000 Europeans, [Novembre et al. \(2008\)](#) showed that a two-dimensional summary of genome-wide variation closely mirrors a map of Europe. A related study by [Lao et al. \(2008\)](#) reported a similar correlation between genetic and geographic structure. The implication for the Hunter Variant Hypothesis is direct. People from neighboring regions of Northern and Northwestern Europe share not just a few pigmentation variants but a broad genetic background. That background includes many of the variants that shape facial form.

This shared background is the population-genetic soil in which recurrence grows. Without invoking any special mechanism, one would expect two unrelated people of, say, English and Danish ancestry to share more appearance-related alleles than two people drawn from very distant populations. That alone makes partial resemblance more likely within the region.

Why this archetype is so visible

Three further factors make the brunette Northern European appearance unusually *visible*, quite apart from its frequency.

- **Diaspora.** Large-scale migration from Northern and Northwestern Europe since the seventeenth century spread these ancestry profiles across North America, Australia, New Zealand and elsewhere. The archetype is common far beyond its region of origin.
- **Media overrepresentation.** Historically, film, television, advertising and publishing in the English-speaking world have drawn disproportionately on performers and models of Northern and Western European descent. As Chapter 11 discusses, visibility follows its own distribution, and this archetype has long sat near the top of it.
- **Historical imagery.** Archival photography, portraiture and, more recently, colorized historical film from Northern and Western Europe circulate widely online. They add a deep historical layer of "variants" to the contemporary one.

Regional texture within the archetype

The archetype is regional, but the region is not uniform. Within Northern and Northwestern Europe, the component traits vary in frequency in ways that give local versions of the archetype their own texture. MC1R variants, and with them auburn tones, fair skin and freckling, are especially frequent in Ireland and Scotland, so brunettes from the Atlantic fringe often carry warmer undertones. Light eye colors reach their highest frequencies around the Baltic and in Finland and Scandinavia, where a brunette with grey or blue eyes is common. In the Low Countries, northern Germany and England, intermediate eye colors such as green and hazel, and medium brown hair, are widespread.

Facial form also shows gentle regional gradients. These are produced by the same mixture of ancestry, drift and local history that shapes pigmentation, but they are subtler and overlap far more. No facial shape belongs to any one nation. The practical point for the Hunter Variant Hypothesis is that a sample of "Northern European brunettes" is itself heterogeneous. A well-designed study should record the broad ancestry of candidates and controls, so that a cluster of variants from one sub-region can be told apart from a cluster driven by sampling.

Archetype, not boundary

It is important to be precise about what "brunette Northern and Northwestern European archetype" does and does not mean. It names a *region of appearance space* that is densely populated by people of Northern and Northwestern European ancestry. It is not a racial category, a genetic boundary or a claim that all people of that ancestry look alike. There is enormous variation within the region, and the traits involved occur, in various combinations, among people of many other ancestries. The [Variant Phenomenon essay](#) puts the general principle well. Race is a broad and socially consequential classification, not a tidy set of discrete biological boxes, and visual resemblance is not evidence of ancestry.

In brief

The brunette Northern and Northwestern European archetype is common, not rare. Brown hair remains the modal adult hair color across most of the region, and fair skin and light eyes reach their highest frequencies there. Shared deep ancestry from three prehistoric streams, and a genetic map that mirrors geography, mean that unrelated people within the region draw on a common pool of appearance-related variants.

05

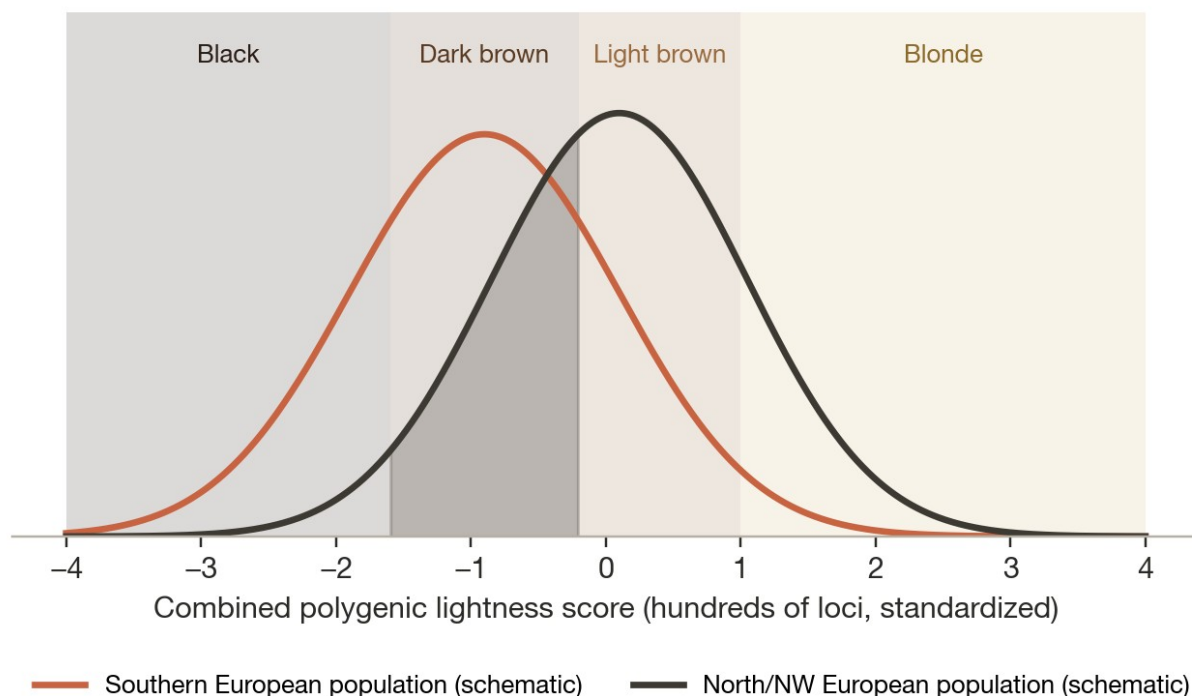
The Genetics of Appearance I: Pigmentation

Color is polygenic

The color of a person's hair, eyes and skin is one of the most heritable traits humans have, and one of the best understood genetically. It is not controlled by a single "brunette gene." Pigmentation is *polygenic*: it is shaped by many genetic variants, each nudging the type and amount of melanin that pigment cells produce. The hair-color GWAS by [Hysi et al. \(2018\)](#) identified more than 120 loci in European-ancestry samples. The UK Biobank analysis by [Morgan et al. \(2018\)](#) found that blond hair alone is associated with over 200 variants and described a *continuum* running from black through dark and light brown to blond.

That continuum is the key to understanding the brunette archetype genetically. Each person inherits a combination of lightening and darkening alleles. The sum of their effects places them somewhere along a scale of pigment production. Medium-to-dark brown hair corresponds to the broad middle of that scale. That is precisely where most people in Northern and Northwestern European populations land, because the frequencies of lightening and darkening variants there are intermediate.

FIGURE 8 *Hair color as a polygenic continuum: schematic distributions for two European populations*



Source: H Heuristics schematic, informed by Morgan et al. (2018) and Hysi et al. (2018). Curves are illustrative, not fitted to data.

The principal pigmentation loci

Although hundreds of variants contribute, a handful of genes have outsized effects. The table below summarizes the loci most relevant to the coloring of the Hunter archetype.

TABLE 5 *Major pigmentation loci relevant to the brunette Northern European archetype*

Gene / region	Principal role	Relevance to the archetype
HERC2 / OCA2	Regulates OCA2 expression; the key switch for blue versus brown eyes	A regulatory variant in HERC2 largely explains blue eye color in Europeans; also influences hair and skin lightness
MC1R	Switches pigment cells between dark eumelanin and red-yellow pheomelanin	Loss-of-function variants cause red hair; carriers often have light brown or auburn tones and fair skin
SLC45A2	Melanosome transporter	Derived allele near-fixed in Northern Europe; contributes to lighter skin and hair
SLC24A5	Ion exchanger in melanosomes	Light-skin allele near-universal across Europe
TYR	Rate-limiting enzyme in melanin synthesis	Variants affect eye and skin lightness and freckling
IRF4	Transcription factor regulating TYR	Associated with hair color, freckling and sun sensitivity
SLC24A4, KITLG, TPCN2	Various melanocyte functions	Smaller effects on hair and eye color shades

Two of these deserve attention. The first is the HERC2 region. [Eiberg et al. \(2008\)](#) and [Sturm et al. \(2008\)](#) independently showed that a single-nucleotide variant in an intron of HERC2 acts as a regulatory switch. It reduces expression of the neighboring OCA2 gene and so lowers melanin in the iris. That one variant explains much of the blue-versus-brown eye difference in Europeans. Because eye color is controlled partly separately from hair color, a person can easily inherit light eyes alongside brown hair. That is exactly the combination that characterizes many brunette Northern Europeans.

The second is MC1R. Most famous as the "red hair gene," [MC1R](#) has many variants. Carrying a single copy is common in Northern and Northwestern Europe, especially in the British Isles, and typically produces warmer brown tones, fair skin that burns easily and a tendency to freckle, without full red hair. Many brunettes in the region carry such variants, which adds subtle warmth and diversity to the brown spectrum.

Independent assortment and the brunette-with-light-eyes combination

Because the genes for hair, eye and skin color sit on different chromosomes or in different regions of the same chromosome, they are inherited largely independently. Different combinations therefore appear routinely within a single population, and even within a single family. A sibling group in Northern Europe might include a blond, blue-eyed child, a brown-haired, green-eyed child and a dark-haired, brown-eyed child. The brunette-with-light-eyes combination at the center of the Hunter archetype is one of the most common of these combinations in the region. It needs no explanation beyond the local allele frequencies.

Why the pigmentation of Northern Europe is so diverse

Europe, and Northern Europe in particular, has an unusually wide range of hair and eye colors by world standards. Population geneticists still debate why. Leading explanations include relaxed constraint on pigmentation at high latitude, positive selection for lighter skin to support vitamin D synthesis under low ultraviolet radiation, and the admixture of ancestral populations with different pigmentation profiles. A more speculative proposal by [Frost \(2006\)](#) invokes frequency-dependent sexual selection. Whatever the mix of causes, the result is a palette of coloring that is both diverse and, in its middle range, very common. The brunette archetype sits in that common middle.

Predicting color from DNA

The pigmentation genetics of Europeans is now well enough understood that hair and eye color can be predicted from DNA with useful accuracy. Forensic [DNA phenotyping](#) systems use a few dozen markers to estimate the probabilities of blue, intermediate and brown eyes, and of blond, brown, red and black hair. These tools are most accurate at the extremes, such as blue versus brown eyes or red versus non-red hair, and less accurate for the intermediate shades. Those intermediate shades include the medium-to-dark brown range that characterizes the Hunter archetype. That fact is a useful reminder that "brunette" is a broad, genetically heterogeneous category: many different allele combinations land in the same visible place.

Coloring across the life course

Pigmentation is not fixed at birth. One of the best-known features of European hair color is that it often *darkens* through childhood and adolescence. Many children who are blond at age three have light or medium brown hair as adults, because melanin production in the hair follicle increases with age and hormonal development. Eye color, too, can change during the first years of life, as melanin accumulates in the iris. In later adulthood, hair loses pigment and greys, which progressively erodes one of the archetype's main cues.

This has two consequences for the hypothesis. First, the brunette archetype is partly an adult phenomenon. Some adult Hunter variants would not have resembled the reference as children, and the reference himself shows a lighter, warmer brown in the adolescent image than in the adult one. Second, comparisons across age groups will understate or overstate resemblance unless age is controlled. A teenager with the reference's adolescent coloring may grow into an adult who resembles the adult reference, or may not. Age-matched comparison is therefore built into the protocol in Chapter 13, and the developmental images in Chapter 02 allow resemblance to be tracked across life stages rather than at a single point.

In brief

The coloring of the Hunter archetype, medium-to-dark brown hair with fair skin and light-to-intermediate eyes, is produced by hundreds of variants of small effect plus a few major loci such as *HERC2/OCA2* and *MC1R*. Because these genes assort independently and their variants are common in Northern Europe, this combination is one of the region's most frequent. Recurrence of the coloring is expected, not surprising.

06

The Genetics of Appearance II: Facial Shape

From color to form

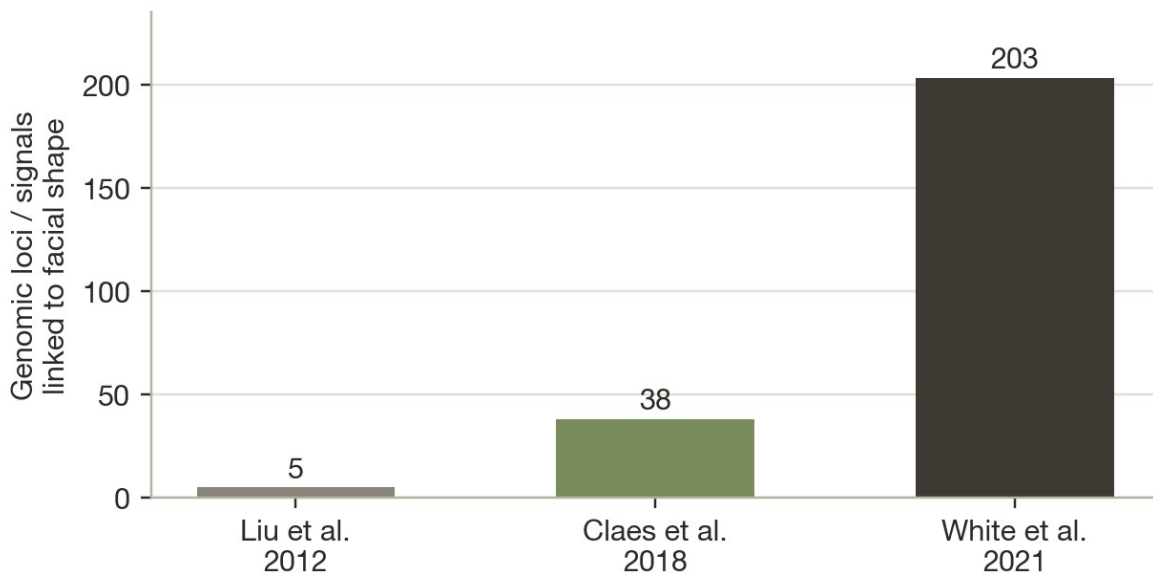
If coloring were the whole story, the Hunter archetype would describe tens of millions of people, and the hypothesis would be trivial. What gives a resemblance its specificity is facial *shape*: the length of the face, the angle of the jaw, the height of the forehead, the set of the brows and the profile of the nose. The genetics of facial shape is younger and more complex than the genetics of pigmentation, but it has advanced rapidly over the last fifteen years.

Facial shape is highly heritable

Twin and family studies show that facial shape is strongly heritable. Identical twins, who share almost all their DNA, look far more alike than fraternal twins, who share about half. Using automated three-dimensional facial phenotyping, [Tsagkrasoulis et al. \(2017\)](#) produced "heritability maps" of the face. They found that heritability is high for many regions, including the nose, the brow ridge and the area around the eyes, with somewhat lower values in regions more affected by soft tissue and environment. In short, the architecture that defines the Hunter archetype is largely, though not entirely, written in the genome.

The growing map of face genes

Genome-wide association studies (GWAS) have identified a steadily increasing number of genetic loci linked to normal facial variation. An early study by [Liu et al. \(2012\)](#) found five loci affecting facial morphology in Europeans. Using a data-driven method that partitions the face into hierarchical segments, [Claes et al. \(2018\)](#) identified 38 loci. By 2021, a multi-ancestry study by [White et al.](#) had mapped 203 genomic signals associated with facial shape. Many lie near genes known to control craniofacial development.

FIGURE 9 Growth in the number of genomic loci linked to normal facial shape variation

Source: Liu et al. (2012); Claes et al. (2018); White et al. (2021).

The picture that emerges closely mirrors that of pigmentation. Facial shape is polygenic. Each locus has a small effect on a specific region of the face, such as the width of the nasal bridge, the projection of the chin or the curvature of the brow. Together they produce the overall form. Many of the genes involved are developmental regulators, such as members of the PAX, SOX and TBX families. They orchestrate how the cells of the embryonic face migrate, proliferate and fuse.

Faces evolved to be distinctive

There is an intriguing evolutionary dimension to all of this. [Sheehan and Nachman \(2014\)](#) found that human facial traits are more variable, and less correlated with one another, than other body measurements. They also found that the genomic regions associated with facial variation carry unusually high levels of genetic diversity. They argued that human faces have evolved, at least in part, *to signal individual identity*. A species that lives in large, complex social groups benefits from faces that are easy to tell apart.

This finding cuts two ways for the Hunter Variant Hypothesis. On one hand, it explains why exact doppelgängers are so rare. Evolution has worked to make faces individual. On the other hand, the relatively *independent* variation of facial features means that partial combinations can recur freely. A person can share the reference's forehead and jaw without sharing his nose, because those features are governed by largely separate genetic modules. That is the genetic basis of the hybrid Hunter.

Convergent recipes

A crucial implication of polygenic architecture is *convergence*. Because each facial feature is influenced by many variants of small effect, two people can arrive at a similar jaw angle or facial length through quite different combinations of alleles. Genetically, they are not "the same," yet visually they converge. This is the second founding proposition of the hypothesis made concrete. It also explains why look-alikes often share *some* genetic variants, as the 2022 study discussed in Chapter 10 found, without being closely related. There are many routes to a similar face, but the routes overlap more than chance would predict.

From DNA to face, and back

The genetics of facial shape is now detailed enough that researchers have begun to test whether a face can be matched to its DNA. [Sero et al. \(2019\)](#) developed face-to-DNA classifiers that use genetic data to pick out the correct face from a set of candidates at better than chance. The accuracy is far from that of fingerprinting, and the technology raises serious ethical questions (Chapter 14). Its scientific significance is clear, however. The specific facial architecture that makes the Hunter archetype recognizable is, to a measurable degree, encoded in heritable variation. That variation is shared across the population in which the archetype is common.

The limits of genetic determinism

It would be a mistake to conclude that a face is fixed by its genome. Environmental and developmental factors contribute meaningfully: nutrition, hormones, injuries, orthodontics, body weight, sun exposure and, above all, age. As the developmental comparison in Chapter 02 shows, the same individual can change considerably between adolescence and adulthood. Soft-tissue features such as the cheeks and lips change across the lifespan, and grooming alters appearance from day to day. Genetic architecture sets the range. Development and life history choose the point within it.

In brief

Facial shape is highly heritable and polygenic, with more than 200 genomic signals now identified. Because facial features vary relatively independently and many allele combinations can yield similar forms, partial resemblance, the hybrid Hunter, is a natural consequence of human facial genetics. Faces evolved to be distinctive overall while leaving room for recurring partial combinations.

07

Biological Constraints: Why Faces Converge

A bounded space

Human appearance is highly variable, but that variation is not unlimited. This line from the [core thesis](#) captures a central idea of developmental biology. All human faces are built by the same developmental program, from the same embryonic tissues, under the same physical and functional constraints. The result is a space of possible faces that is vast but bounded, and far from uniformly filled.

The developmental blueprint

Every human face develops from the same embryonic structures: the frontonasal prominence and the paired maxillary and mandibular processes. These grow, meet and fuse between roughly the fourth and tenth weeks of gestation. The timing, rate and extent of their growth, guided by signaling pathways such as SHH, BMP, FGF and WNT, determine the eventual proportions of the face. Because these pathways are deeply conserved, variation among people is produced by modest adjustments to a shared program rather than by different programs. That is why all human faces share a basic configuration, and why the differences between them are differences of degree.

Functional constraint

The face is also a working organ system. It must house the eyes, nose and mouth. It must accommodate the airway, support chewing and protect the brain. These functions impose further limits. The eyes must sit at a certain distance apart for binocular vision. The jaw must be strong enough for chewing, and the nasal cavity must allow airflow. Faces that depart too far from functional norms are rare, because the developmental program rarely produces them and because they may carry costs. Most human faces therefore cluster within a central band of proportions.

Integration and modularity

Developmental biologists describe facial traits as partly *integrated* and partly *modular*. Integrated traits tend to vary together. A longer face, for example, tends to come with a longer nose and a taller mid-face, because they share developmental influences. Modular traits vary more independently, such as eye color and nasal shape. Integration produces convergence: once one feature moves in a direction, related features tend to follow, so faces with similar overall proportions tend to resemble each other in several correlated ways at once. Modularity produces the hybrid pattern, in which some features match and others do not.

The elongated-to-angular face of the Hunter archetype is a good example of integration. A vertically oriented facial skeleton typically comes with a higher forehead, a longer mid-face and a more defined mandibular angle in adulthood. These are not independent coincidences. They tend to travel together. A candidate who shares the overall vertical proportion of the reference will often share several of these correlated features as well, which makes the resemblance seem greater than the sum of its parts.

Averageness and the dense center of face space

The bounded, centrally clustered distribution of faces has a perceptual consequence explored further in Chapter 09. Faces near the center of the distribution, those close to the population average, are by definition more numerous. In a famous study, [Langlois and Roggman \(1990\)](#) showed that composite faces made by averaging many individual faces are rated as more attractive than most of the individual faces that went into them. Averageness is also linked to familiarity: average faces tend to feel more familiar and to be confused with one another more often.

The idea of a composite is old. In the 1870s, [Francis Galton](#) superimposed photographs to create "composite portraits," hoping to reveal typical faces of particular groups. His underlying goals were deeply flawed and tied to eugenics, and the method has been rightly criticized. But the technique itself captured a real phenomenon. Faces drawn from a population share a common structure, and blending them reveals that structure. In a sense, a recurring archetype is a population's composite showing through in many individuals.

Convergence through shared environment and culture

Biological constraints are reinforced by shared environmental and cultural factors. People in the same region share similar diets, similar childhood health and similar patterns of sun exposure, all of which affect facial development and skin. They also share grooming norms, fashionable hairstyles and expressions shaped by cultural display rules. A generation of young men in the same cultural context may adopt similar hairstyles, beard styles and poses for photographs. That adds a convergent presentational layer on top of the biological one. This is one reason why Type IV, or Stylistic, variants can cluster in particular eras and subcultures.

What constraint does not imply

None of this means that faces are interchangeable or that a small number of "types" can describe humanity. Constraint makes the space bounded, not small. Even with shared developmental rules, the number of distinguishable faces is enormous, which is why most people can tell thousands of individuals apart. The point is narrower and more precise. Constraint, integration and averageness make some *regions* of face space densely populated and make partial resemblance within those regions common. The Hunter archetype, a combination of common coloring with a recognizable but not extreme facial architecture, sits in such a region.

In brief

All human faces are built by a shared, conserved developmental program under functional constraints, so the space of possible faces is bounded and clustered. Integrated traits travel together and create convergent resemblances. Modular traits vary independently and create hybrid ones. The Hunter archetype sits in a populated region of this space, where recurrence is structurally favored.

08

Statistical Probability: The Mathematics of Recurrence

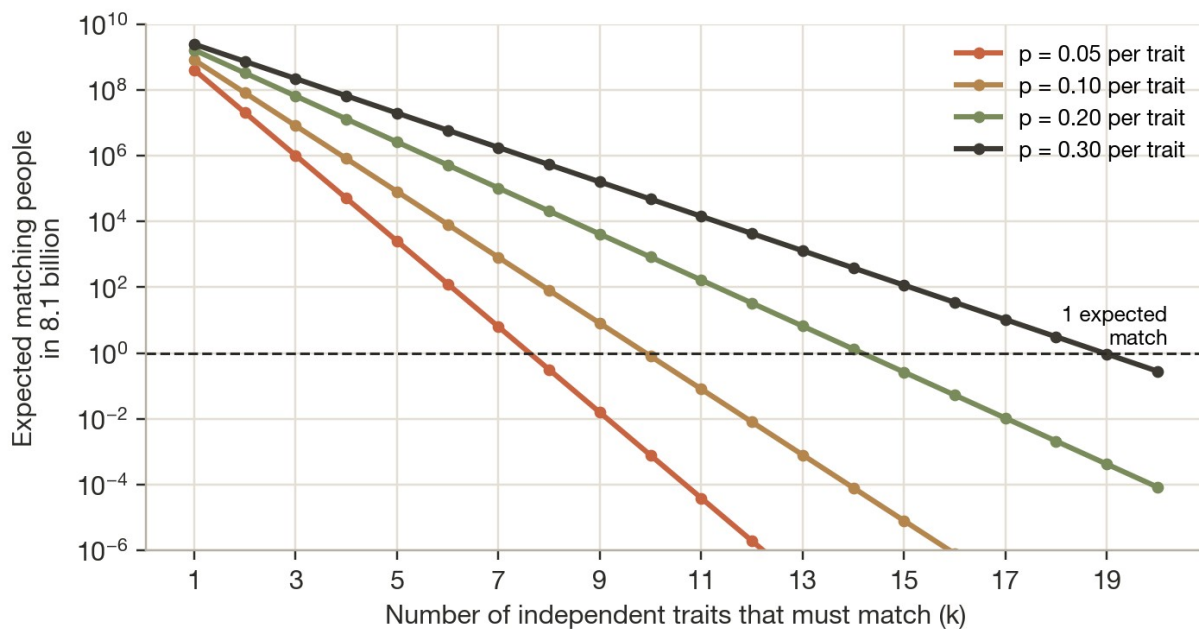
Eight billion draws from a shared pool

The deposited preprint frames the central puzzle in its title: [why similar faces recur in a world of eight billion people](#). The answer starts with arithmetic. With roughly 8.1 billion people alive today, and many billions more preserved in photographs, film and portraiture, even very improbable combinations of traits will occur many times over. The question is not *whether* recurrence happens but *how often*, and at what level of resemblance.

A simple combinatorial model

Suppose a resemblance requires a match on k independent traits, and that each trait has a probability p of matching the reference in a randomly chosen person. The probability that a random person matches on all k traits is then p to the power k , and the expected number of matches in a population of size N is $N \times p^k$. The model is deliberately simplified. Real traits are neither fully independent nor equally probable. But it captures the essential dynamics.

FIGURE 10 Expected number of matching individuals among 8.1 billion people, by number of required traits



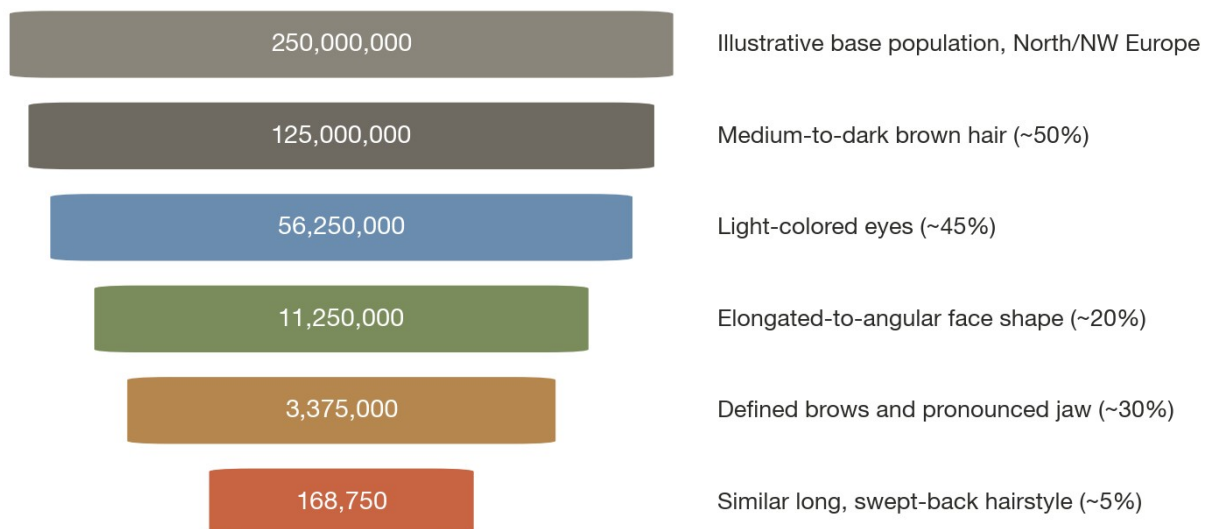
Source: H Heuristics model. Assumes independent traits with equal match probability p .

The figure shows two lessons. First, recurrence is extremely sensitive to the number of traits that must match. For common traits, those with a match probability of 0.3, a ten-trait match is expected in about 48,000 people worldwide. For traits with a match probability of 0.1, the same ten-trait match is expected in fewer than one person. Second, partial matches are vastly more common than complete ones. Every trait dropped from the definition multiplies the expected count by a factor of three to twenty. This is the mathematical engine behind the dominance of the hybrid Hunter. A match on four or five common traits is not rare at all.

The funnel of the Hunter archetype

The same logic can be applied directly to the Hunter archetype. The figure below begins with an illustrative base population of 250 million people of Northern and Northwestern European ancestry. It then applies successive filters for the defining traits, each with a rough, deliberately conservative frequency.

FIGURE 11 *Illustrative funnel: how many people might share the defining traits of the Hunter archetype*



Source: H Heuristics illustrative model. Frequencies are rough assumptions chosen for exposition, treated as independent; they are not measured prevalences.

Even after six successive filters, including a filter for the distinctive swept-back hairstyle that only about one in twenty people might share, the model leaves roughly 170,000 people. If the filters are relaxed to the coloring and broad facial proportions that drive Type II and Type III resemblance, the pool expands into the millions. If the diaspora populations of North America and Australasia are added, it expands again. The exact numbers depend on assumptions that only a real prevalence study could replace. The order of magnitude, however, is robust. Even a strict definition of the Hunter archetype describes a large population of potential variants.

How unique is a face?

A related question has been examined directly in forensic anthropology. Using the ANSUR database of 3,982 individuals, [Lucas and Henneberg \(2015\)](#) searched for people whose faces were metrically identical across eight facial measurements. They found that no two people matched exactly once seven or more

measurements were combined, and they proposed the term "singular" rather than "unique." The [University of Adelaide's press release](#) popularized the result as odds of about one in a trillion for an exact eight-measurement match. That interpretation was later challenged on statistical grounds in a [published reply by Meester, Preneel and Wenmackers \(2019\)](#).

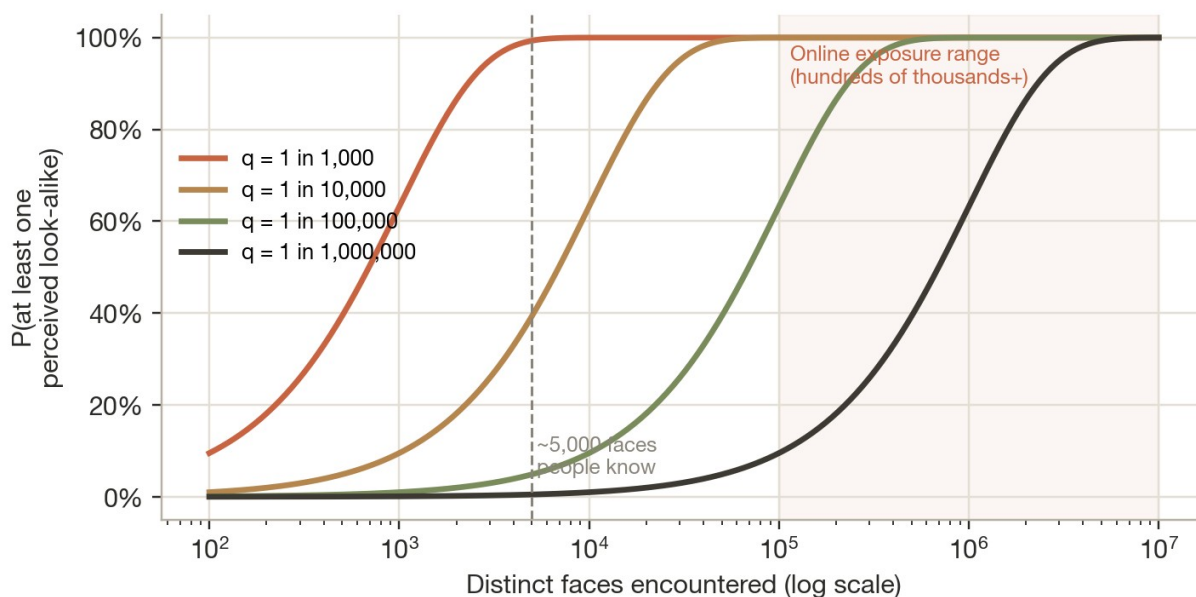
For the Hunter Variant Hypothesis, the important point is the distinction between *metric identity* and *perceived resemblance*. Two faces that match within millimeters on every measurement are vanishingly rare. Two faces that a human observer would describe as "looking alike" are common. Observers tolerate large metric differences when the salient cues align (Chapter 09). The hypothesis concerns the second phenomenon, not the first.

The birthday problem for faces

The classic [birthday problem](#) shows that in a group of just 23 people, the chance that two share a birthday exceeds 50 percent. The reason is that the number of *pairs* grows much faster than the number of people. The same logic applies to look-alikes. [Jenkins, Dowsett and Burton \(2018\)](#) estimated that people know roughly 5,000 faces on average. Among 5,000 known faces there are about 12.5 million distinct pairs. If only one pair in a million counts as a striking look-alike pair, a typical person's acquaintance network would still contain a dozen such pairs.

From the observer's side, what matters is the chance of encountering at least one face that resembles a *given* reference, such as a friend, a partner or oneself. If a fraction q of faces resemble the reference closely enough to be noticed, the probability of seeing at least one after encountering n faces is $1 - (1 - q)^n$.

FIGURE 12 Probability of encountering at least one perceived look-alike of a given reference, by exposure



Source: H Heuristics model. q denotes the fraction of faces that resemble the reference closely enough to be noticed.

For a resemblance threshold of one in 10,000, an observer who knows 5,000 faces has about a 39 percent chance of knowing at least one look-alike of a given person. An observer who scrolls through 100,000 faces online faces near-certainty. Even at a much stricter threshold of one in a million, exposure to ten million faces, which is quite plausible over years of intensive social-media use, makes at least one

encounter overwhelmingly likely. The experience of repeatedly "seeing Hunter everywhere" is therefore exactly what probability predicts for an observer with high exposure.

Non-independence makes recurrence more likely, not less

The simple models above assume independent traits, but real traits are correlated. In Chapter 07 we saw that integrated facial features travel together, and Chapter 04 showed that pigmentation traits co-occur within populations. Correlation *increases* the expected number of multi-trait matches compared with the independent model. A person who matches on one integrated feature is more likely than chance to match on its correlates. The independence assumption therefore makes the combinatorial estimates conservative. The real pool of partial matches is probably larger.

A worked example: from coloring to hybrid

The difference between the archetype's components is easiest to see in a worked example. Suppose an observer in a Northern European city passes 1,000 adult strangers in a week. Under the rough frequencies used in the funnel above, about 500 will have brown hair and about 225 of those will have light eyes. Coloring alone therefore offers more than two hundred people a week who match the reference on the most visible cue. Of these, perhaps 45 will have a broadly elongated face, and perhaps 13 will add defined brows and a strong jaw. Fewer than one in a typical week will also share the distinctive hairstyle.

This simple arithmetic explains a pattern that the H Heuristics essays describe qualitatively. Feature-specific, Type III resemblances based on coloring are encountered constantly and quickly forgotten. Hybrid, Type II resemblances, which combine coloring with facial proportion, occur a handful of times a week in a dense urban setting and are memorable when noticed. Strong, Type I resemblances that also share the presentational signature are rare in person but become common online, where the observer's weekly sample runs into the tens or hundreds of thousands. The spectrum of variant types is, in part, a spectrum of base rates.

Base rates and the meaning of a match

The worked example also clarifies why base rates matter for interpretation. If a large share of a population shares the archetype's coloring, then a coloring match carries little information about anything beyond coloring. A match is informative only to the extent that it is *unexpected* given the base rate. This is why the testing protocol compares nominated candidates not with random faces but with matched brunette controls. A candidate who resembles the reference more than other brunettes of the same background tells us something about the specific archetype. A candidate who resembles him only as much as any brunette does tells us about the population.

In brief

In a population of eight billion, recurrence is a statistical certainty at the level of partial resemblance. Each additional trait required for a match shrinks the pool steeply, which is why exact doppelgängers are rare and hybrid variants common. An illustrative six-filter model of the Hunter archetype still describes about 170,000 people in Northern and Northwestern Europe alone. With high exposure, encountering a look-alike of any given reference becomes nearly inevitable.

09

Human Perception: How Resemblance Is Recognized

Resemblance happens in the observer

Probability explains how many potential variants exist. Perception explains which of them are *noticed*. The [original essay](#) recognizes that resemblance is partly a perceptual judgment, shaped by the observer's familiarity and expectations. This chapter examines the cognitive science of face recognition to show precisely how that judgment is made, and why it so readily produces the experience of variants.

The specialized face system

Humans have a dedicated neural system for processing faces. Neuroimaging work by [Kanwisher, McDermott and Chun \(1997\)](#) identified a region of the fusiform gyrus, the [fusiform face area](#), that responds more strongly to faces than to other objects. The influential cognitive model of [Bruce and Young \(1986\)](#) describes recognition as a sequence. Structural encoding of a face is followed by comparison with stored "face recognition units," then by activation of identity information. A sense of familiarity can arise at the comparison stage *before* identity is retrieved. That is exactly the "I know that face from somewhere" experience that often precedes the recognition of a variant.

Holistic and configural processing

Faces are not processed as a list of independent parts. They are processed *holistically*, as integrated wholes, and *configurally*, with attention to the spatial relationships between features. A review by [Maurer, Le Grand and Mondloch \(2002\)](#) distinguishes three components: sensitivity to the basic first-order arrangement of two eyes above a nose above a mouth; holistic binding of features into a gestalt; and sensitivity to second-order spacing differences between features.

Holistic processing explains the Type V, or *Archetypal*, variant. If the brain encodes faces as wholes, two faces can share an overall gestalt even when no single feature matches precisely. The impression of resemblance is real, but it cannot be broken down into features. Holistic processing also explains why hybrid resemblances feel stronger than they "should": a partial feature match can be enough to activate the stored whole.

Familiar and unfamiliar faces

One of the most robust findings in face research is the asymmetry between familiar and unfamiliar faces. People recognize familiar faces with remarkable reliability across changes in lighting, age, angle and expression. They are far worse at matching *unfamiliar* faces, even when the two images are presented side by side. In controlled matching tasks, error rates of 10 to 30 percent are common.

A striking demonstration comes from [Jenkins et al. \(2011\)](#). Participants unfamiliar with two Dutch celebrities were asked to sort forty photographs into piles, one per person. They typically created many

more than two piles, around seven on average. Participants familiar with the two people sorted them almost perfectly. The variability *within* one person's photographs was, for unfamiliar viewers, as large as the variability *between* people. As [Burton \(2013\)](#) argued, this within-person variability is central to understanding face recognition.

This asymmetry has a precise implication for the hypothesis. Observers who know the reference well, such as friends, family and the subject himself, have a rich, variable representation of his face. They may notice resemblances in strangers that match *some* version of the face they know: Hunter with a beard, Hunter as a teenager, Hunter in a particular light. Meanwhile, unfamiliar observers comparing photographs will make many errors in both directions. Any test of the hypothesis must account for rater familiarity.

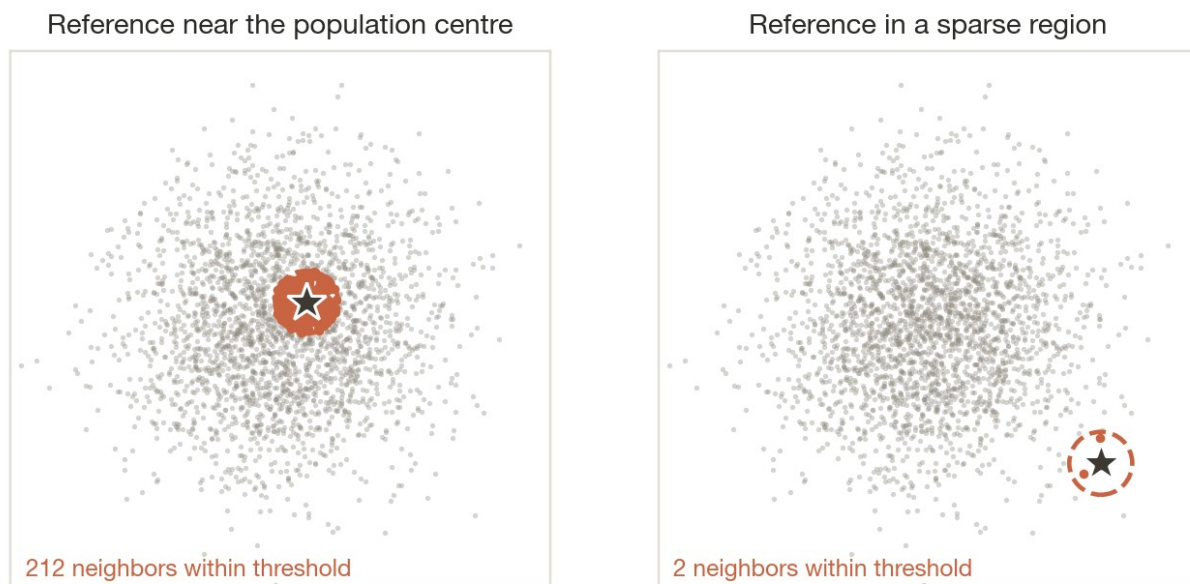
Cue salience: hair, outline and brows

When people process unfamiliar faces, they rely disproportionately on *external* features such as the hairline, hairstyle and face outline, together with the high-contrast brow and eye region. Internal features become more important as faces grow familiar. For the Hunter archetype, this means that the long, swept-back brown hair, the vertical face outline and the straight, defined brows are exactly the cues that strangers will weight most heavily. A candidate who matches on these external cues can be perceived as a strong variant by unfamiliar observers even if the internal features differ considerably.

The face-space framework

[Valentine's \(1991\)](#) *face-space* model offers a unifying picture. Each face is represented as a point in a multidimensional space whose dimensions correspond to the features that distinguish faces. The population average sits at the center, and most faces cluster around it. Distinctive faces lie in sparse outer regions, while typical faces lie in the crowded center.

FIGURE 13 *Face-space illustration: the number of near neighbors depends on where the reference sits*



Source: H Heuristics simulation after Valentine (1991). 2,600 simulated faces; dashed circle marks a fixed similarity threshold.

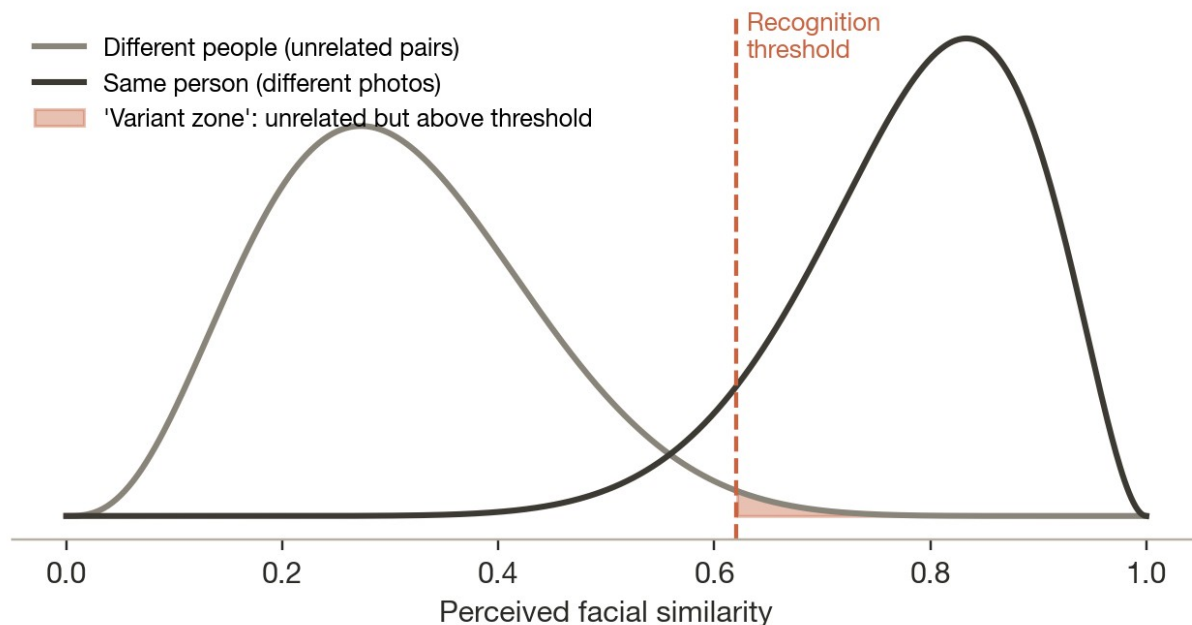
The model predicts that a reference face near the populated center of its population's face space will have many neighbors within any fixed similarity threshold, and therefore many potential variants. A reference

face in a sparse region will have few. As Chapters 04 to 07 showed, the Hunter archetype combines the region's modal coloring with a recognizable but not extreme facial architecture. Within the face space of Northern and Northwestern European populations, it is likely to sit in a well-populated region, closer to the left panel than the right. This is a testable prediction. Algorithmic embedding (Chapter 10) can estimate how dense the reference's neighborhood actually is.

Signal detection: the variant zone

Recognition can be described in terms of signal detection theory. For any pair of images, the observer forms an internal sense of similarity and decides "same" or "different" depending on whether it crosses a threshold. Similarity values for pairs showing the same person and for pairs of different people form two overlapping distributions.

FIGURE 14 *Signal detection view of resemblance: unrelated pairs that cross the recognition threshold form the "variant zone"*



Source: H Heuristics illustration. Distributions are schematic.

The right-hand tail of the "different people" distribution, the pairs of unrelated people whose similarity exceeds the observer's threshold, is the *variant zone*. Every variant ever reported comes from this tail. Its size depends on the observer's threshold, which shifts with expectations, attention and motivation. An observer primed to look for Hunter variants effectively lowers the threshold, and the variant zone expands.

Biases that shape resemblance judgments

Several well-documented cognitive biases affect how resemblance is perceived and remembered.

- **Confirmation bias.** People seek and remember evidence that supports what they already believe, as reviewed by [Nickerson \(1998\)](#). Hits, the variants found, are memorable. Misses, the thousands of brunettes who did not resemble the reference, are forgotten.

- **The frequency illusion.** Once a pattern has been noticed, it seems to appear everywhere. This [frequency illusion](#) reflects selective attention, not a change in frequency.
- **The cross-race and cross-group effect.** Observers are generally better at distinguishing faces from groups they have more experience with, as shown in a meta-analysis by [Meissner and Brigham \(2001\)](#). Observers with less exposure to Northern European faces may perceive more resemblance among them, while observers from within the group may perceive less.
- **Pareidolia and the drive to find faces.** The brain readily detects faces in ambiguous stimuli, a phenomenon known as [pareidolia](#). The same pattern-seeking tendency operates at the level of identity.
- **Individual differences.** Face recognition ability varies enormously. At one extreme are [super-recognizers](#), and at the other are people with [prosopagnosia](#). The same pair of faces may look identical to one observer and quite different to another.

Implications for documenting variants

These findings translate into practical rules for anyone documenting variants, whether in an informal archive or a formal study.

- **Record the viewing conditions.** Note whether the image is a photograph, a video still, a colorized archive image or an artwork, and whether lighting, angle and resolution allow facial structure to be judged at all.
- **Separate cues from impressions.** Record which specific cues drive the resemblance, such as hair, brows or outline, before recording an overall impression. This guards against holistic impressions being reported as structural matches.
- **Log misses as well as hits.** A note of how many brunette faces were seen without a resemblance being noticed turns an anecdote into the start of a base rate.
- **Note who noticed.** Record whether the resemblance was noticed by the reference subject, by someone who knows him, or by a stranger, since familiarity changes what observers see.
- **Wait before deciding.** Revisiting a nominated image after a delay, without the reference images side by side, helps separate a robust resemblance from a momentary impression.

Taken together, these rules shift the archive from a collection of striking moments toward a dataset that can support inference. They also put the perceptual science of this chapter directly to work.

In brief

Resemblance is constructed by a specialized face system that processes faces holistically and, for unfamiliar faces, leans heavily on external cues such as hair and face outline. Variants come from the "variant zone": unrelated pairs whose similarity exceeds an observer's threshold. That threshold moves with expectation, familiarity and bias, so any serious test must control for the observer as carefully as for the faces.

10

Machines and Look-Alikes: Biometrics and Genetic Evidence

Algorithms as observers

Human resemblance judgments are rich but noisy. Modern face-recognition algorithms provide a second, independent kind of observer. They are less flexible than humans and carry their own biases, but they are consistent, scalable and quantitative. Their use in *doppelgänger* research has also produced one of the most interesting findings relevant to the Hunter Variant Hypothesis.

How face-recognition systems represent faces

Contemporary systems use deep neural networks to map each face image to a point, or *embedding*, in a high-dimensional space, typically with 128 to 512 dimensions. The network is trained so that images of the same person land close together and images of different people land far apart. The influential [FaceNet](#) system (Schroff, Kalenichenko and Philbin, 2015) introduced a training method based on triplets of images that has shaped much of the field. Similarity between two faces is then simply a distance between their embeddings.

This is, in effect, a machine-learned face space: a practical, quantitative version of Valentine's theoretical model. It offers exactly what the Hunter Variant Hypothesis needs. Each candidate variant can be placed at a measurable distance from the reference, and the *density* of the reference's neighborhood can be measured directly.

Look-alikes as a biometric problem

For biometric security, look-alikes are a failure mode called a *false match*. The U.S. National Institute of Standards and Technology evaluates face-recognition algorithms in its [Face Recognition Vendor Test program](#). It has documented that false-match rates rise for pairs of people who are demographically similar and are especially high for twins. In other words, the most capable algorithms in the world find some unrelated people hard to tell apart. Algorithms, like humans, have a variant zone.

A study by [Phillips et al. \(2018\)](#) compared professional forensic facial examiners, super-recognizers, untrained participants and leading algorithms on difficult face-matching tasks. The top algorithms performed in the range of the best human experts, and the most accurate results came from *combining* human and machine judgments. This supports a hybrid design for testing the Hunter Variant Hypothesis, in which blind human ratings and algorithmic similarity scores are collected side by side.

The 2022 look-alike genetics study

The most directly relevant empirical study is [Joshi et al. \(2022\)](#), published in *Cell Reports* by a team led by Manel Esteller. The researchers recruited pairs of unrelated look-alikes photographed for an art project by a Canadian photographer. They first tested the pairs with three independent face-recognition

algorithms, then asked whether the most algorithmically similar pairs also resembled each other genetically and biologically.

Their key findings were these. A subset of the pairs was rated as highly similar by all three algorithms, with similarity scores comparable to those of identical twins. Among these algorithmically confirmed look-alikes, many shared significantly more genetic variants than random pairs from the same population, including variants near genes associated with facial features. By contrast, the pairs differed in epigenetic profiles (DNA methylation) and in their microbiomes. The authors also reported that the look-alikes tended to share some physical and lifestyle characteristics, such as height and weight, in questionnaire data.

The authors interpreted these results as evidence that some unrelated look-alikes share part of a genetic "recipe" for facial appearance, a common set of variants that produces similar faces, and that the recurrence of such combinations is to be expected in a large, growing population. The study had real limitations. The sample was small and mostly of European ancestry, the pairs were self-selected, and the genetic analysis cannot identify specific causal variants.

What the study means for the hypothesis

The Joshi study offers the strongest empirical support yet for the *convergence* and *recurrence* propositions of the Hunter Variant Hypothesis, with an important qualification. It shows that striking resemblance between unrelated people is not purely a perceptual illusion. At least in some cases, it reflects genuine overlap in appearance-related genetic variation. That overlap is exactly what the population-genetic picture of Chapters 04 to 06 predicts for people drawn from populations with shared deep ancestry, as most of the study participants were.

The qualification is that the study concerns strong, algorithmically confirmed look-alikes, which correspond to Type I in the variant spectrum. It says little about the hybrid, feature-specific and stylistic variants that make up most of the phenomenon. Nor does it support any idea of a single "variant gene." The shared variants were many, of small effect, and widely distributed. They were common building blocks, not a lineage marker. The [Variant Phenomenon essay](#) frames the finding appropriately as a scientific foothold rather than a confirmation of the full framework.

Machine limits and biases

Algorithmic similarity is a useful instrument, not ground truth. Systems vary in their training data, and performance can differ across demographic groups. Embeddings encode features that matter for identity verification, which may not coincide with the features humans use for resemblance. In particular, many systems are trained to be *invariant* to hairstyle, which is one of the most salient cues for human observers and central to the Hunter archetype. A candidate who looks very like the reference to people because of a shared hairstyle may score as dissimilar to an algorithm, and vice versa. Divergence between human and machine judgments is therefore informative in itself. It can help tell Type IV stylistic variants apart from structural ones.

What a reference-centred embedding study would show

Applied to the Hunter Variant Hypothesis, algorithmic tools allow three measurements that human raters cannot easily provide. The first is a *neighborhood profile*. For a large, consented image set, the distance from the reference embedding to every other face can be computed and sorted. The shape of that distribution, and in particular how many faces fall within the distances typical of same-person photo

pairs, tells us how crowded the reference's neighborhood is compared with the neighborhoods of other people in the same sample.

The second is a *stability check*. Embedding distances can be recomputed for reference images taken at different ages, with different hairstyles and with the hair covered. A candidate who remains close to the reference across all versions is a structural variant. A candidate who is close only to the styled version is a stylistic one. A candidate who is close only to the adolescent image points to developmental resemblance.

The third is *cross-model agreement*. Different face-recognition models are trained on different data and make different errors. A candidate who is ranked close to the reference by several independent models is much less likely to be an artifact of one system's quirks. Taken together, these three measurements would provide a reproducible, machine-readable description of the Hunter variant neighborhood. Human ratings could then be compared against that description, and the gap between the two would itself be a finding about perception.

In brief

Face-recognition algorithms provide a consistent, quantitative second observer that measures similarity as distance in a learned face space. A 2022 *Cell Reports* study found that unrelated look-alikes confirmed by such algorithms often shared more genetic variants than random pairs. That gives genuine support to the idea that strong resemblance can reflect a shared polygenic recipe. It does not support any single "variant gene."

11

The Internet as Variant Amplifier and the Pareto Distribution of Visibility

A new field of observation

For most of human history, a person's sample of faces was limited to their community: perhaps a few thousand people over a lifetime. Photography, film and television expanded that sample. The internet has expanded it by orders of magnitude. A single afternoon of scrolling short-form video can expose a viewer to more faces than a medieval villager would have seen in a lifetime. The [original essay](#) describes the internet as both a *discovery environment* for variants and a *source of sampling bias*. This chapter examines both roles.

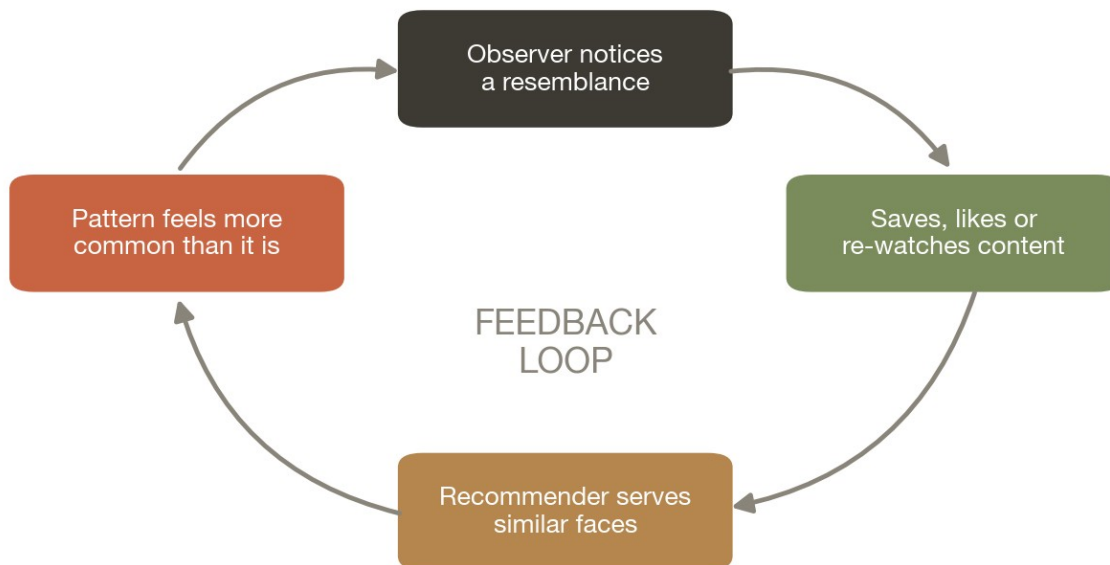
Exposure scales recurrence

As Chapter 08 showed, the probability of encountering a look-alike of a given reference rises steeply with the number of faces encountered. The internet's main effect is to push every observer far along that curve. Over years of use, exposure to hundreds of thousands or millions of faces makes encounters with variants not just possible but routine. The informal *Variants* album in Chapter 03 shows this. Its contents came largely from social-media screens, streaming platforms, online video and digitized archives.

The feedback loop

The internet does not just expose observers to more faces. It shapes *which* faces they see, in response to what they engage with. Recommendation systems learn from engagement. A viewer who pauses on, likes or saves content featuring people who resemble the reference sends a signal that may lead the system to serve more content with similar visual features. Over time this creates a feedback loop.

FIGURE 15 *The variant amplifier: engagement-driven recommendation can inflate perceived prevalence*



Source: H Heuristics analysis.

The result is that the observer's personal stream of faces becomes enriched for the archetype they are looking for. The pattern really does appear more often *in their feed*, but this tells us about the feed, not the population. As the [Variant Phenomenon essay](#) puts it, the internet is a powerful place to generate hypotheses but not, by itself, a representative population survey.

Three different claims

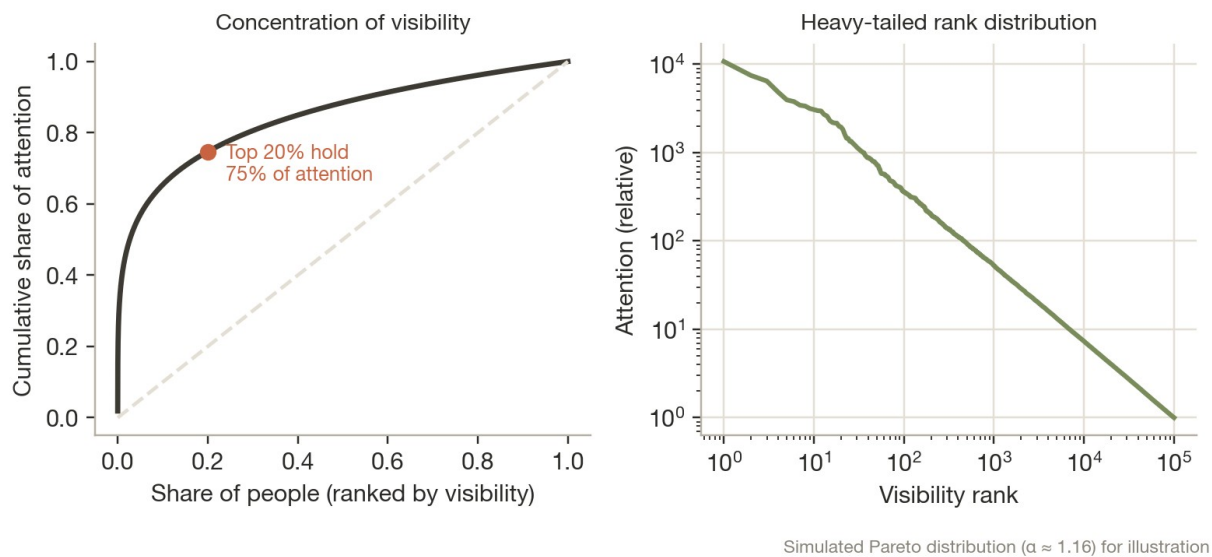
The essay draws a distinction that is central to evaluating any variant observation. A variant may be:

- **common in the population**, a claim about biological and demographic prevalence;
- **common among public figures**, a claim about who becomes visible; or
- **common in the content an observer sees**, a claim about a specific, filtered stream of media.

These three claims require different evidence, and confusing them is the most common error in reasoning about variants. Online observations directly support only the third. The first requires population sampling. The second requires comparing public figures with matched, less-visible people.

Alpha variants and the Pareto distribution of visibility

The [Variant Phenomenon essay](#) introduces the provisional term *alpha variant* for an archetype that is disproportionately represented among highly visible people, such as actors, musicians, athletes, entrepreneurs, commentators and creators. The essay is explicit that "alpha" refers to prominence, not biological superiority. The concept draws on a well-documented property of attention: it is extremely unequally distributed.

FIGURE 16 *The Pareto distribution of visibility: a small share of people receive most of the attention*

Source: H Heuristics simulation (Pareto $\alpha \approx 1.16$, which yields the classic 80/20 split).

The **Pareto principle**, the observation that roughly 80 percent of an outcome often comes from 20 percent of causes, is an empirical regularity rather than a universal law, as the essay notes. Attention online, however, does follow heavy-tailed distributions. Network models of **preferential attachment** (Barabási and Albert, 1999) show how such distributions arise when new attention flows preferentially to those who already have it, a dynamic sociologists call the **Matthew effect**.

Why some archetypes become alpha variants

If visibility is heavily concentrated, then the appearances of the few people at the top of the distribution are seen far more often than their population frequency would suggest. Several mechanisms could make particular archetypes overrepresented there:

- **Casting and beauty norms.** Entertainment and advertising have historically favored certain appearances, and these choices reflect and reinforce cultural standards of attractiveness. Research on facial attractiveness, reviewed by Little, Jones and DeBruine (2011), points to averageness, symmetry and sexual dimorphism as consistent factors, alongside cultural variation.
- **Trait inferences from faces.** People rapidly infer traits such as competence and trustworthiness from faces. Todorov et al. (2005) found that inferences of competence made from brief exposure to candidates' faces predicted U.S. congressional election outcomes at well above chance. Such judgments may give some facial types an advantage in reaching visibility, regardless of actual ability.
- **Demographic composition of media industries.** As noted in Chapter 04, English-language media have historically drawn heavily on people of Northern and Western European descent, which mechanically raises the visibility of archetypes common in those populations.
- **Network effects.** Once a look becomes associated with success in a field, imitation through styling and grooming spreads it further, creating clusters of Type IV variants around prominent figures.

Interpreting alpha status carefully

The alpha-variant concept generates testable questions. Is the brunette Northern European archetype overrepresented among, say, leading actors relative to its share of the relevant population? Is the specific Hunter archetype overrepresented within that group? The essay rightly insists that any such comparison must use matched controls and account for selection effects. It also insists that prominence says nothing about capability or worth. Appearance does not predict personality or destiny, and the framework explicitly rejects any ranking of groups. Alpha status is a fact about attention, and attention is shaped by culture, economics and chance as much as by faces.

In brief

The internet multiplies each observer's exposure to faces and, through engagement-driven recommendation, enriches their stream for the archetypes they notice. Visibility follows a heavy-tailed, Pareto-like distribution, so archetypes common among prominent people, the *alpha variants*, are seen far more often than their population frequency implies. Perceived frequency, prominence and true prevalence are three different claims requiring three kinds of evidence.

12

The Variant Matrix and the "Variant Gene" Metaphor

From one archetype to many

The Hunter Variant Hypothesis is one instance of a broader H Heuristics idea. The [companion essay](#) introduces the *variant matrix*, defined as the broader landscape of recurring visual archetypes, and positions the Hunter variant as one region within it that H Heuristics has chosen to study. Related essays explore other regions, including [The European Variant Matrix](#), which looks at history's recurring human types, and [The Recurring Human Face and the Lars Variant](#), which proposes a companion archetype. A general treatment appears in the H Heuristics report on [The Human Variant Hypothesis](#).

Mapping the matrix

A simple way to picture the variant matrix is as a grid whose axes are major appearance dimensions. The figure below crosses hair color with broad facial form, two of the most salient cues in face perception. Each cell is a region of appearance space that may contain its own recurring archetypes. The Hunter variant sits at the intersection of medium-to-dark brown hair and an elongated-to-angular facial form. The cell shading is a schematic indication of how populated each region might be among people of Northern and Northwestern European ancestry.

FIGURE 17 *A simplified two-dimensional slice of the variant matrix*

	Blonde	Light brown	Medium-dark brown	Black	Red
Round / soft	...	...	...	...	...
Oval / balanced	Lars-type (companion)	...	...	...	...
Elongated / angular	...	...	HUNTER VARIANT	...	...
Broad / square	...	...	...	...	...

Source: H Heuristics schematic. Cell shading is illustrative; the placement of the companion "Lars-type" archetype is indicative only.

A real variant matrix would have many more dimensions, including eye color, skin tone, nasal form, brow shape, age and presentation. It would also be estimated from data rather than drawn by hand. In the language of Chapter 10, the matrix is a coarse, human-readable summary of the structure of face space.

Machine embeddings could reveal its true geometry, including which regions are dense, which are sparse and where natural clusters form.

Variants across all populations

The variant matrix is not a European construct. As the [Variant Phenomenon essay](#) emphasizes, every broad human population, whether African, East Asian, South Asian, Southeast Asian, Indigenous, Middle Eastern, European, Latin American or of mixed ancestry, may contain its own recurring archetypes. The same mechanisms of polygenic inheritance, developmental constraint, statistical recurrence and perception operate everywhere. Labels such as "African variant" or "East Asian variant" are only meaningful as descriptions of patterns observed in clearly defined samples. They are not claims about discrete biological types. This report's focus on the Brunette Northern and Northwestern European archetype is one slice of a general phenomenon.

The "Hunter variant gene": a metaphor, not a mechanism

The original essay uses the phrase "Hunter variant gene" and immediately clarifies that it is a metaphor. The hypothesis currently offers no genetic evidence for such a mechanism. That clarification deserves emphasis, because the phrase could easily be misread.

In a literal sense, there is no Hunter variant gene. As Chapters 05 and 06 showed, the traits that make up the archetype are each influenced by hundreds of genetic variants of small effect, spread across the genome. No single gene, or small set of genes, produces "a Hunter." Two people who resemble the reference will typically share some of the relevant variants, and the 2022 look-alike study suggests that strong look-alikes share more than chance. But they will reach their resemblance through overlapping, not identical, genetic combinations.

The metaphor is still useful when read correctly. It points to a real phenomenon: a *recipe* of common variants that, when combined, tends to produce the archetype. A better technical image might be a *polygenic neighborhood*, a region of genetic variation whose members tend to land in the same region of face space. That image keeps the intuition while staying faithful to the science.

TABLE 6 *Metaphorical versus literal readings of key H Heuristics terms*

Term	Metaphorical / framework meaning	What it does <i>not</i> mean
Hunter variant	A person whose appearance recurs part of the reference archetype	A relative, descendant or biological subtype
Variant gene	Shorthand for a shared polygenic recipe that tends to produce the archetype	A single gene or inherited marker of the archetype
Variant matrix	The landscape of recurring visual archetypes across humanity	A fixed set of human "types" or a racial classification
Alpha variant	An archetype overrepresented among highly visible people	Biological or moral superiority of any appearance
25–50% likeness	Descriptive shorthand for partial resemblance	A measured statistic or genetic percentage

Building the matrix from data

The variant matrix drawn above is a sketch. A data-driven version could be built in three steps. First, a large, consented and demographically documented set of face images would be encoded with a face-recognition model, placing every face in a common embedding space. Second, unsupervised clustering methods would be used to look for regions of unusually high density, natural "neighborhoods" that recur across independent samples. Third, human raters would be shown representative faces from each neighborhood and asked to describe what the faces share. That step gives each region an interpretable label, such as "medium-brown hair, elongated face, defined brow," instead of a number.

This approach has an important advantage over hand-built typologies. It lets the data decide whether recurring archetypes exist at all, and how many there are, rather than assuming a fixed set in advance. If the embedding space turns out to be a smooth continuum without distinct clusters, the variant matrix would be better described as a *map* of a continuous landscape than as a grid of types. That would itself be a meaningful result. It would also give the Hunter variant a precise location: the neighborhood of face space around the reference, with a measurable density and boundary.

A data-driven matrix would also make it possible to place the H Heuristics companion archetypes, such as the Lars variant, relative to the Hunter variant. Distances between neighborhoods would show which archetypes are close cousins in appearance and which are genuinely distinct, and would show how those relationships hold across populations.

Why the distinction matters

Keeping metaphor and mechanism apart is not a technicality. Throughout the twentieth century, claims linking appearance to ancestry, character or worth were used to justify serious harms. The face-based typologies of physiognomy and the eugenic projects associated with figures such as Galton are cautionary examples. The H Heuristics essays explicitly disavow these readings. The framework does not claim that appearance predicts personality or destiny, that any variant is more capable, or that groups should be ranked. Using precise language is how that commitment is made operational in research.

In brief

The Hunter variant is one region of a broader *variant matrix* that spans all human populations. The "Hunter variant gene" is a metaphor for a shared polygenic recipe, a neighborhood of genetic variation that tends to produce the archetype. It is not a literal gene or a lineage. Precise language keeps the framework scientific and guards against misuse.

13

Testing the Hypothesis: A Research Design

From impression to evidence

The H Heuristics essays repeatedly describe the hypothesis as exploratory: a hypothesis "in search of evidence." This chapter takes that seriously. It sets out a research design that could test the main claims, building on the testing sections of the [original essay](#) and the [Variant Phenomenon essay](#), and on standard practice in face-perception research.

Testable claims

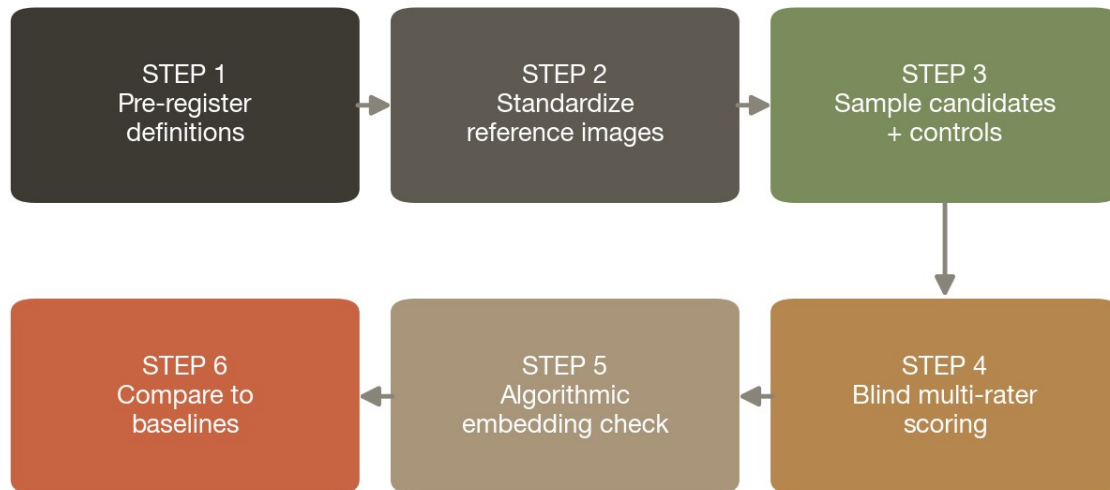
The hypothesis can be broken down into specific claims, each with a distinct test.

TABLE 7 *Core claims of the Hunter Variant Hypothesis and how each could be tested*

Claim	Operational test	Supporting evidence would be
H1. Recognizable recurrence	Blind raters score candidates vs. reference	Candidate ratings above matched-control ratings
H2. Partial resemblance dominates	Distribution of multi-dimension scores	Most above-threshold cases show uneven (hybrid) profiles
H3. Dense neighborhood	Algorithmic embedding density around reference	Reference has more near neighbors than typical faces in the sample
H4. Archetype regionality	Neighbor density by ancestry sample	Near neighbors concentrated in North/NW European samples
H5. Alpha overrepresentation	Prevalence among public figures vs. matched non-public sample	Higher prevalence among public figures after controls
H6. Amplification	Compare feed-derived vs. random-sample prevalence	Feed-derived rate exceeds random-sample rate

A six-step protocol

FIGURE 18 Six-step protocol for testing the Hunter Variant Hypothesis



Source: H Heuristics research design.

Step 1 — Pre-register definitions

Before any data are collected, the archetype, the scoring dimensions, the thresholds for each variant type and the analysis plan should be fixed and **publicly pre-registered**. This turns the informal "25 to 50 percent likeness" shorthand into a pre-specified score, for example a weighted composite of the five dimension ratings, with category boundaries set in advance. Pre-registration is the single most important protection against the confirmation bias described in Chapter 09.

Step 2 — Standardize reference images

A set of reference images should be captured under controlled conditions: frontal and three-quarter views, neutral and smiling expressions, consistent lighting, and both with the hair styled as usual and with the hair pulled back or covered. Hair-covered images let raters and algorithms separate structural from stylistic resemblance. The adolescent reference supports developmental comparisons.

Step 3 — Sample candidates and controls

Three groups of images are needed. *Nominated candidates* are people proposed as variants by observers, including entries from the existing archive. *Matched brunette controls* are people with similar coloring, age and sex but selected without regard to resemblance. *Random controls* are drawn from a broader population. The matched-control group is essential. It answers the critical question of whether candidates resemble the reference *more than other brunettes of the same background do*. Without it, the study would only confirm that the reference is a brunette Northern European.

Step 4 — Blind multi-rater scoring

Raters who do not know which images are candidates and which are controls should score each image against the reference. They should rate each of the five dimensions separately on a fixed scale and give an overall resemblance score and a confidence rating last. Raters should include both people unfamiliar with

the reference and people familiar with him, analyzed separately. Inter-rater reliability should be reported using intraclass correlation coefficients, following guidance such as [Koo and Li \(2016\)](#).

Step 5 — Algorithmic embedding check

Each image should also be encoded with one or more established face-recognition models, and its embedding distance from the reference computed. Algorithmic scores provide a consistent comparison and allow the density of the reference's neighborhood to be estimated. They should be computed for both hair-visible and hair-covered images, given the hairstyle-invariance of many systems (Chapter 10).

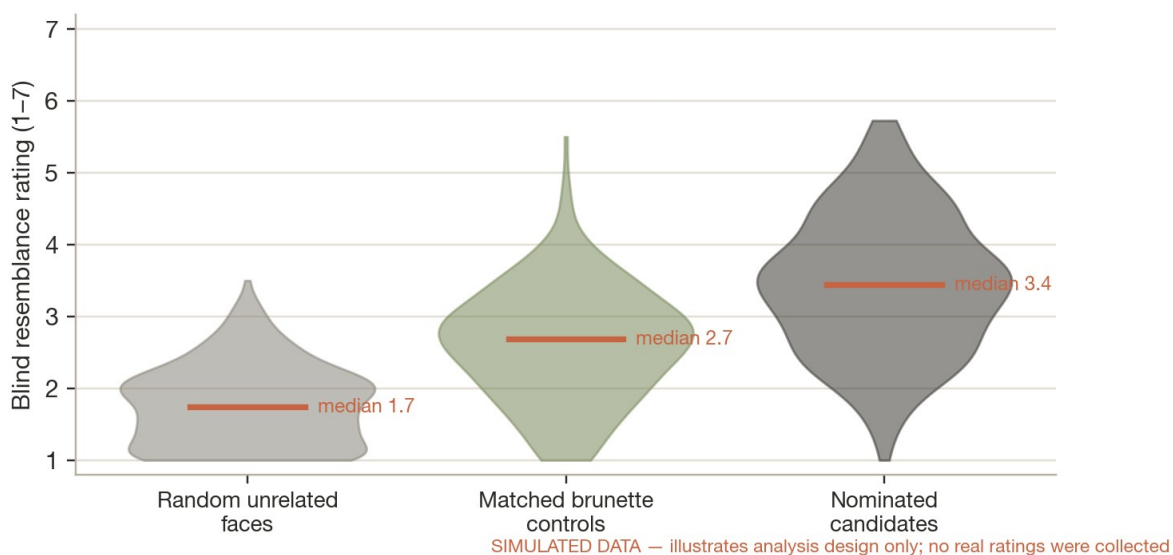
Step 6 — Compare to baselines

The main analysis compares the three groups. If candidates are rated as substantially more similar than matched controls, H1 is supported. If above-threshold cases mostly show uneven multi-dimensional profiles, H2 is supported. Agreement and disagreement between human and machine scores help classify cases as structural or stylistic.

What results might look like

To make the design concrete, the figure below shows *simulated* data of the kind such a study might produce. It does not report real results, and no ratings have been collected. It illustrates the analysis: nominated candidates would need to score clearly above matched brunette controls, not merely above random faces, for the hypothesis to gain support.

FIGURE 19 Illustrative analysis using simulated ratings for three comparison groups



Source: H Heuristics simulation. SIMULATED DATA for design illustration only; no real ratings were collected.

What a null result would mean

A well-designed test must be able to fail. Several null results are possible, and each would be informative.

- **Candidates no more similar than matched controls.** This would suggest that perceived variants are mostly *generic* brunette Northern Europeans, and that the impression of a specific recurring Hunter archetype reflects coloring plus observer expectation.

- **Strong human ratings but weak algorithmic similarity.** This would indicate that resemblance is driven mainly by presentation, especially hairstyle, which points to Type IV dominance.
- **Reference neighborhood no denser than average.** This would weaken the claim that the archetype sits in an especially populated region of face space.

As the Variant Phenomenon essay notes, a null result would still be useful. It would refine the framework, clarify which components of resemblance are real and which are perceptual, and strengthen the credibility of H Heuristics' research program.

Prevalence and prominence studies

Testing H4 and H5 requires larger samples. For regionality, embedding density around the reference can be compared across image samples of different ancestries, using properly consented research datasets. For alpha overrepresentation, the prevalence of the archetype among a defined set of public figures, such as leading film actors from a given period, can be compared with prevalence in a demographically matched sample of non-public individuals. Selection effects, including casting norms and the demographic composition of the industry, must be modeled explicitly.

Sample size and statistical power

A pilot study does not need to be large to be informative, but it should be planned with statistical power in mind. If nominated candidates truly differ from matched brunette controls by a moderate effect, roughly half a standard deviation on the composite resemblance score, then about 64 images per group would give 80 percent power to detect the difference at conventional significance levels in a simple two-group comparison. Each image would be rated by at least ten independent raters. A smaller effect, such as a quarter of a standard deviation, would need about four times as many images. Because ratings are nested within raters and images, the main analysis should use mixed-effects models with random effects for both, rather than simple averages.

Power analysis also clarifies what the pilot can and cannot establish. A pilot with a few dozen images per group can test whether nominated candidates are rated as more similar than controls (H1) and can estimate the distribution of profiles (H2). It cannot estimate population prevalence (H4) or test alpha overrepresentation (H5), which need far larger and carefully constructed samples. Planning these as separate phases avoids over-reading a pilot.

In brief

The hypothesis can be tested with a pre-registered, six-step protocol: standardized reference images, nominated candidates compared against *matched brunette controls*, blind multi-dimensional human ratings, algorithmic embedding checks and baseline comparisons. The design specifies in advance what support and failure look like, and treats a null result as a finding.

14

Ethics, Privacy and the Limits of the Framework

Research on faces is research on people

Every face in a variant archive belongs to a person. That simple fact shapes the ethics of the Hunter Variant Hypothesis. The [companion essay](#) already raises concerns about platform rules, copyright, privacy and unwanted attention, and those concerns should be built into the research program from the start.

Privacy and biometric data

Facial images, and especially algorithmic face embeddings, are increasingly treated in law as sensitive biometric data. In the European Union, [Article 9 of the GDPR](#) classes biometric data used to uniquely identify a person as a special category requiring explicit legal grounds for processing. In the United States, laws such as Illinois's [Biometric Information Privacy Act](#) impose consent and retention requirements on the collection of face geometry. A research program that computes embeddings of people found online must account for these frameworks, and in many cases it should restrict algorithmic analysis to consenting participants or licensed research datasets.

Consent and the public-figure question

The reference subject has consented to being studied. The people identified as variants have not. Several principles follow:

- **No identification of private individuals.** Variant archives should not name, tag or locate private individuals, or invite others to do so. This report follows that rule. The individuals in the informal *Variants* album are not identified.
- **Care with public figures.** Public figures have a reduced expectation of privacy about their public image, but they do not consent to being described as someone else's "variant." Commentary should be respectful and avoid implying any personal connection.
- **No contact or pursuit.** The research should never encourage contacting, following or seeking out people because of their appearance.
- **Consent for primary data.** New images for systematic study should come from participants who have given informed consent, ideally under independent ethical review.

Copyright and image rights

Screenshots from films, streaming platforms, broadcasts and social media are usually the copyright of their creators or distributors. Informal, non-commercial commentary may fall within fair use or fair dealing in some jurisdictions. Systematic publication, and especially commercial use, requires more care.

A sustainable archive should store links and metadata in preference to reproducing images, use licensed or public-domain material where possible, and respond promptly to takedown requests.

Avoiding typological misuse

The history of face-based typology is a serious warning. Physiognomy, race science and eugenics all claimed that faces revealed character, capability or worth, and all were used to justify discrimination and violence. Any framework that groups people by appearance carries a risk of being read in that light, whatever its intentions. The H Heuristics framework guards against this explicitly. The [Variant Phenomenon essay](#) states that it does not claim a fixed number of human types, that appearance predicts personality or destiny, that a prominent variant is more capable, or that groups should be ranked or character inferred from faces. This report adopts those boundaries without qualification.

Limits of the framework

Beyond ethics, the framework has real epistemic limits.

- **Observer dependence.** Resemblance is partly in the eye of the beholder. Even with blind ratings, results will depend on the rater population.
- **Image dependence.** Photographs vary in lighting, angle, lens and processing. Historical, colorized and artistic images are especially unreliable as evidence of underlying facial structure.
- **Selection.** Any archive assembled by noticing resemblances is biased toward hits. Only systematic sampling can estimate prevalence.
- **Genetic inference.** Visual resemblance is not evidence of ancestry or relatedness. Genetic claims require genetic data, collected with consent.
- **Commercial interest.** The companion essay openly acknowledges a commercial interest in the phenomenon. That is legitimate, but it creates an incentive to overstate findings. Independent review, pre-registration and open data are the best counterweights.

TABLE 8 *Ethical safeguards for the Hunter Variant research program*

Risk	Safeguard
Identification of private individuals	Never name, tag or locate; blur or omit where needed
Biometric processing without consent	Embeddings only for consenting participants or licensed datasets
Copyright infringement	Link rather than reproduce; licensed sources; takedown process
Typological misreading	Explicit disclaimers; no trait or worth inferences; precise terminology
Overstatement for commercial reasons	Pre-registration; independent review; publish null results

Responsible audience participation

Because the hypothesis invites audiences to contribute observations, the research program also needs simple rules for participation. Contributors should be asked to submit links or descriptions rather than uploading images of private individuals. They should be told plainly that submissions will not be used to identify, locate or contact anyone, and that content showing private individuals without consent will be

removed. Public figures may be discussed, but in a respectful register that describes resemblance without speculating about the person's private life, ancestry or character.

Participatory experiments, such as blind rating games, should use only consented or licensed images, explain how responses will be used, and allow participants to opt out. Collected responses should be stored without personal identifiers wherever possible. These rules protect the people in the images, protect contributors from unintended consequences, and keep the archive usable as research data. A short, public participation policy, published alongside the variant archive, is the simplest way to make the rules visible and enforceable.

In brief

Research on variants is research on real people who have not consented to being studied. The program should avoid identifying private individuals, treat face embeddings as sensitive biometric data, respect image rights and keep firm boundaries against any inference of character or worth from appearance. Pre-registration and openness protect both the participants and the credibility of the work.

15

Implications for H Heuristics and a Research Roadmap

Why this hypothesis matters to H Heuristics

H Heuristics describes itself on its [About page](#) as a founder-led, multidisciplinary organization working at the intersection of research, media, business and civic discourse. Its intellectual program is concerned with recurring structures and the underlying architecture of reality. The Hunter Variant Hypothesis fits that program unusually well. It takes a recurring structure everyone has experienced, the recurring human face, and subjects it to the same structural, statistical and systems-level analysis that H Heuristics applies to economics, technology and geopolitics.

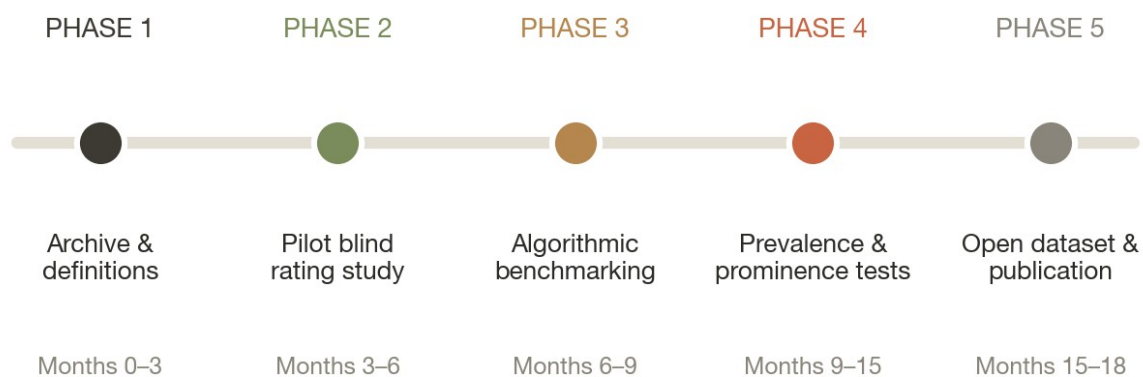
The hypothesis also has a distinctive public dimension. As the [companion essay](#) notes, the reference identity creates a natural link between the founder, the organization and its audience. Audiences can contribute observations, test their own perceptions and follow the development of the research. Few research programs are so immediately participatory.

Building a credible program

The opportunity comes with a responsibility. The value of the Hunter Variant Hypothesis to H Heuristics depends on its credibility. A program that overstates its evidence would damage the organization's wider research reputation. A program that tests its claims openly, reports what it finds and treats null results as findings would strengthen it. The essays themselves set this standard by speaking of capitalizing on recognition *without overstating the evidence*.

An eighteen-month roadmap

FIGURE 20 *Proposed research roadmap for the Hunter Variant program*



Source: H Heuristics research plan.

Phase 1 — Archive and definitions (months 0–3)

Formalize the variant archive as a structured database recording source, date, image conditions, visible characteristics, provisional type and coder confidence, using links rather than stored images where rights are unclear. Finalize and pre-register the operational definitions and scoring instrument (Appendix B).

Phase 2 — Pilot blind rating study (months 3–6)

Capture standardized reference images. Recruit a pilot panel of raters, both familiar and unfamiliar with the reference. Score a sample of nominated candidates alongside matched brunette controls and random controls. Estimate inter-rater reliability and refine the instrument.

Phase 3 — Algorithmic benchmarking (months 6–9)

Using consenting participants or licensed datasets, compute embedding distances between the reference and the pilot images for multiple face-recognition models, with hair visible and covered. Compare machine and human scores to separate structural from stylistic resemblance.

Phase 4 — Prevalence and prominence tests (months 9–15)

Estimate the density of the reference's neighborhood within defined image samples, and compare it with typical faces. Test alpha overrepresentation by comparing a defined group of public figures with demographically matched non-public samples, modeling selection effects explicitly.

Phase 5 — Open dataset and publication (months 15–18)

Publish the protocol, the anonymized results, the analysis code and, where consent allows, the dataset. Deposit a full research article alongside the existing preprints on [figshare](#) and [Zenodo](#), and submit it for peer review.

Content and communication

Alongside the formal research, the hypothesis lends itself to a public communication strategy that is both engaging and responsible:

- **Explainer content** on the science behind recurring faces, covering polygenic inheritance, face space, the birthday problem and the variant zone. It turns a personal phenomenon into accessible science.
- **Participatory experiments** in which audiences rate pairs of faces under blind conditions and see how their own judgments compare. These can be used as pilot data where properly consented.
- **Transparent updates** that report progress through the roadmap, including results that do not support the hypothesis.
- **Extension to the wider variant matrix**, linking the Hunter archetype to the Lars variant, the European Variant Matrix and the population-general Human Variant Hypothesis.

Measures of success

Success for the program should be defined by research quality, not only reach: a pre-registered protocol published; inter-rater reliability at acceptable levels; a completed pilot comparing candidates with matched controls; an open dataset; and at least one peer-reviewed or formally reviewed publication. Audience growth is welcome, but it is a by-product of credible work rather than its measure.

Risks to the program

Three risks deserve explicit management. The first is *confirmation drift*: over time, a research team that wants the hypothesis to succeed may relax its definitions or reinterpret ambiguous results. Pre-registration, version-controlled definitions and an external reviewer who can see the protocol before results are known are the standard defenses. The second is *audience pressure*. Public engagement rewards striking examples, and a feed of dramatic look-alikes can crowd out the slower work of controls and baselines. Keeping public content clearly separate from research outputs helps protect both. The third is *scope creep*. The variant matrix offers almost unlimited extensions, from the Lars variant to population-general archetypes. Completing the Hunter study to publication before widening the program will produce a stronger foundation than starting many studies at once.

In brief

The Hunter Variant Hypothesis gives H Heuristics a distinctive, participatory research program that fits its interest in recurring structures. Its value depends on credibility. An eighteen-month roadmap, from archive and pre-registration through blind rating, algorithmic benchmarking and prevalence testing to open publication, turns the hypothesis into a research asset.

CONCLUSION

Conclusion: A Recurring Human Form

The Hunter Variant Hypothesis starts from a human experience: seeing someone we know in the face of a stranger. It turns that experience into a structured question. Why does a particular appearance, here the brunette Northern and Northwestern European archetype represented by Hunter Hughes, appear again and again among people who have no connection to one another?

This report has argued that the answer lies in four interacting layers. **Genetics** supplies a shared pool of appearance-related variants. In Northern and Northwestern Europe, that pool makes medium-to-dark brown hair, fair skin and light or intermediate eyes some of the most common combinations. Facial shape is equally polygenic, and its features vary independently enough for partial combinations to recur freely. **Biological constraint** bounds and clusters the space of possible faces, so that many people share a region of it and integrated features travel together. **Statistical probability** makes recurrence inevitable at the level of partial resemblance in a population of eight billion. Even a strict definition of the archetype describes a large population of potential variants. And **human perception**, holistic, cue-driven and threshold-based, turns some of those potential variants into experienced ones, with the internet multiplying exposure and amplifying what we notice.

Two conclusions follow. First, the phenomenon the hypothesis describes is real but not mysterious. It needs no hidden mechanism, and the hypothesis has never proposed one. The best empirical evidence available, the 2022 look-alike genetics study, suggests that strong resemblances can reflect a shared polygenic recipe. That is what the population genetics of the region would predict. Second, the most interesting claims of the hypothesis are still open: that a *specific* recurring archetype can be recognized reliably, that partial resemblance dominates, that the reference sits in a dense neighborhood of face space and that some archetypes are overrepresented among the highly visible. These are testable, and this report has set out how they could be tested.

Human individuality and recurring form are not in tension. Every person is unique in genome, history and identity. Every face is also built from a shared vocabulary, by a shared developmental grammar, and read by a shared perceptual system. The Hunter Variant is one sentence that this grammar keeps writing. Recognizing it, measuring it and understanding it is a modest but genuine contribution to the larger H Heuristics project of mapping the recurring structures of reality.

Final Statement

The Hunter Variant Hypothesis is best held as a resemblance hypothesis in search of evidence: grounded in genetics, constrained by biology, driven by probability and shaped by perception. Its claims are clear, its tests are specified, and its boundaries are explicit. What remains is the data.

APPENDIX A

Glossary of Terms

TABLE 9 *Glossary of key terms used in this report*

Term	Definition
Alpha variant	H Heuristics term for an archetype disproportionately represented among highly visible people; refers to prominence only.
Archetype	A recognizable region of appearance space that recurs across unrelated individuals.
Composite resemblance	The holistic impression that a person "reads as" the reference, independent of specific features.
Configural processing	Perception of the spatial relationships between facial features.
Convergence	Arrival at similar appearances through different underlying genetic or developmental routes.
Doppelgänger	Popular term for an unrelated person who closely resembles another; corresponds to Type I.
Embedding	A numerical vector representing a face in a machine-learned similarity space.
Face space	A theoretical multidimensional space in which each face is a point and similarity is distance.
False match	In biometrics, the incorrect acceptance of two different people as the same.
GWAS	Genome-wide association study; a method for finding genetic variants associated with traits.
Holistic processing	Perception of a face as an integrated whole rather than as separate parts.
Hybrid Hunter	A Type II variant sharing multiple, but not all, traits of the reference.
Polygenic	Influenced by many genetic variants, each of small effect.
Pre-registration	Public specification of hypotheses and analysis plans before data collection.
Variant	A working label for a person who recurs part of a recognizable archetype.
Variant matrix	The landscape of recurring visual archetypes across all human populations.
Variant zone	The tail of the unrelated-pairs similarity distribution that exceeds an observer's recognition threshold.

APPENDIX B

Resemblance Scoring Instrument

The instrument below operationalizes the five dimensions for blind rating (Chapter 13). Raters view the reference image set and one candidate image at a time, in randomized order, without knowing whether the candidate is a nominated variant or a control.

TABLE 10 *Hunter Variant Resemblance Scoring Instrument (rater form)*

Item	Prompt to rater	Scale
D1 Hair & coloring	How similar are the hair color, hair texture, skin tone and eye color?	1 (not at all) – 7 (nearly identical)
D2 Facial structure	How similar are the face outline, length, forehead and jaw?	1 – 7
D3 Feature configuration	How similar are the brows, eyes, nose, lips and their spacing?	1 – 7
D4 Presentation	How similar are the hairstyle, facial hair, grooming and expression?	1 – 7
D5 Overall resemblance	Overall, how much does this person resemble the reference? (rate last)	1 – 7
C Confidence	How confident are you in your ratings for this image?	1 (guess) – 5 (certain)
F Familiarity	How well do you know the reference person? (once per session)	None / slight / moderate / well

Scoring. A pre-registered composite score, for example the mean of D1 to D3 weighted toward structure plus D5, determines variant type against pre-specified thresholds. Profiles with high D4 but low D2 and D3 are classified as Type IV (Stylistic). High D5 with low and inconsistent D1 to D4 indicates Type V (Archetypal). Hair-covered image sets are scored separately on D2, D3 and D5.

TABLE 11 *Pre-specified classification rules (illustrative thresholds to be fixed at pre-registration)*

Type	Rule (mean across raters)
Type I · Strong	$D5 \geq 5.5$ and at least three of $D1$ – $D3 \geq 5.0$
Type II · Hybrid	$D5 \geq 4.0$ and two of $D1$ – $D3 \geq 4.5$, with at least one ≤ 3.5
Type III · Feature-specific	One of $D1$ – $D3 \geq 5.0$, others < 4.0
Type IV · Stylistic	$D4 \geq 5.0$, $D2$ and $D3 < 4.0$
Type V · Archetypal	$D5 \geq 4.5$ with no single dimension ≥ 4.5
Not a variant	$D5 < 4.0$ and no dimension ≥ 5.0

SOURCES

References

H HEURISTICS SOURCES

- Hughes, H. (2026). *The Hunter Variant Hypothesis: Recurring Doppelgängers, Hybrid Resemblance, and the Emergence of a Personal Archetype*. H Heuristics. hheuristics.substack.com/p/the-hunter-variant-hypothesis
- Hughes, H. (2026). *Hunter Hughes, H Heuristics, and the Hunter Variant Phenomenon*. H Heuristics. hheuristics.substack.com/p/hunter-hughes-h-heuristics-and-the-6e7
- Hughes, H. (2026). *The Variant (or Doppelgänger) Phenomenon: Recurring Human Archetypes, Alpha Variants, and the Pareto Distribution of Visibility*. H Heuristics. hheuristics.substack.com/p/the-variant-or-doppelganger-phenomenon
- Hughes, H. (2026). *The Hunter Variant Hypothesis: Why Similar Faces Recur in a World of Eight Billion People – Genetics, Biometrics, and the Architecture of Human Appearance*. Preprint, figshare. [doi:10.6084/m9.figshare.34856661](https://doi.org/10.6084/m9.figshare.34856661)
- Hughes, H. (2026). *The Hunter Variant Hypothesis* (original deposit). Zenodo. [doi:10.5281/zenodo.23764959](https://doi.org/10.5281/zenodo.23764959)
- H Heuristics. *The European Variant Matrix*. hheuristics.substack.com/p/the-recurring-cast
- H Heuristics. *The Recurring Human Face and the Lars Variant*. hheuristics.substack.com/p/the-recurring-human-face-and-the
- H Heuristics. *The Human Variant Hypothesis*. researchreports.hheuristics.com

GENETICS AND POPULATION HISTORY

- Claes, P., et al. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nature Genetics*, 50, 414–423. [doi:10.1038/s41588-018-0057-4](https://doi.org/10.1038/s41588-018-0057-4)
- Eiberg, H., et al. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene. *Human Genetics*, 123, 177–187. [doi:10.1007/s00439-007-0460-x](https://doi.org/10.1007/s00439-007-0460-x)
- Frost, P. (2006). European hair and eye color: A case of frequency-dependent sexual selection? *Evolution and Human Behavior*, 27, 85–103. [doi:10.1016/j.evolhumbehav.2005.07.002](https://doi.org/10.1016/j.evolhumbehav.2005.07.002)
- Haak, W., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522, 207–211. [doi:10.1038/nature14317](https://doi.org/10.1038/nature14317)
- Hysi, P. G., et al. (2018). Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nature Genetics*, 50, 652–656. [doi:10.1038/s41588-018-0100-5](https://doi.org/10.1038/s41588-018-0100-5)
- Joshi, R. S., et al. (2022). Look-alike humans identified by facial recognition algorithms show genetic similarities. *Cell Reports*, 40, 111257. [doi:10.1016/j.celrep.2022.111257](https://doi.org/10.1016/j.celrep.2022.111257)
- Lao, O., et al. (2008). Correlation between genetic and geographic structure in Europe. *Current Biology*, 18, 1241–1248. [doi:10.1016/j.cub.2008.07.049](https://doi.org/10.1016/j.cub.2008.07.049)

- Lazaridis, I., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513, 409–413. [doi:10.1038/nature13673](https://doi.org/10.1038/nature13673)
- Liu, F., et al. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics*, 8, e1002932. [doi:10.1371/journal.pgen.1002932](https://doi.org/10.1371/journal.pgen.1002932)
- Mathieson, I., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503. [doi:10.1038/nature16152](https://doi.org/10.1038/nature16152)
- Morgan, M. D., et al. (2018). Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nature Communications*, 9, 5271. [doi:10.1038/s41467-018-07691-z](https://doi.org/10.1038/s41467-018-07691-z)
- Novembre, J., et al. (2008). Genes mirror geography within Europe. *Nature*, 456, 98–101. [doi:10.1038/nature07331](https://doi.org/10.1038/nature07331)
- Sero, D., et al. (2019). Facial recognition from DNA using face-to-DNA classifiers. *Nature Communications*, 10, 2557. [doi:10.1038/s41467-019-10617-y](https://doi.org/10.1038/s41467-019-10617-y)
- Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications*, 5, 4800. [doi:10.1038/ncomms5800](https://doi.org/10.1038/ncomms5800)
- Sturm, R. A., et al. (2008). A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *American Journal of Human Genetics*, 82, 424–431. [doi:10.1016/j.ajhg.2007.11.005](https://doi.org/10.1016/j.ajhg.2007.11.005)
- Tsagkrasoulis, D., et al. (2017). Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific Reports*, 7, 45885. [doi:10.1038/srep45885](https://doi.org/10.1038/srep45885)
- White, J. D., et al. (2021). Insights into the genetic architecture of the human face. *Nature Genetics*, 53, 45–53. [doi:10.1038/s41588-020-00741-7](https://doi.org/10.1038/s41588-020-00741-7)

FACE PERCEPTION, STATISTICS AND BIOMETRICS

- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512. [doi:10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509)
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305–327. [doi:10.1111/j.2044-8295.1986.tb02199.x](https://doi.org/10.1111/j.2044-8295.1986.tb02199.x)
- Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, 66, 1467–1485. [doi:10.1080/17470218.2013.800125](https://doi.org/10.1080/17470218.2013.800125)
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121, 313–323. [doi:10.1016/j.cognition.2011.08.001](https://doi.org/10.1016/j.cognition.2011.08.001)
- Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B*, 285, 20181319. [doi:10.1098/rspb.2018.1319](https://doi.org/10.1098/rspb.2018.1319)
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area. *Journal of Neuroscience*, 17, 4302–4311. [doi:10.1523/JNEUROSCI.17-11-04302.1997](https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997)
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. [doi:10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)

- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115–121. [doi:10.1111/j.1467-9280.1990.tb00079.x](https://doi.org/10.1111/j.1467-9280.1990.tb00079.x)
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society B*, 366, 1638–1659. [doi:10.1098/rstb.2010.0404](https://doi.org/10.1098/rstb.2010.0404)
- Lucas, T., & Henneberg, M. (2015). Are human faces unique? A metric approach to finding single individuals without duplicates in large samples. *Forensic Science International*, 257, 514.e1–514.e6. [doi:10.1016/j.forsciint.2015.09.003](https://doi.org/10.1016/j.forsciint.2015.09.003)
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6, 255–260. [doi:10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4)
- Meester, R., Preneel, B., & Wenmackers, S. (2019). Reply to Lucas & Henneberg: Are human faces unique? *Forensic Science International*, 297, 217–220. research.vu.nl
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7, 3–35. [doi:10.1037/1076-8971.7.1.3](https://doi.org/10.1037/1076-8971.7.1.3)
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220. [doi:10.1037/1089-2680.2.2.175](https://doi.org/10.1037/1089-2680.2.2.175)
- Phillips, P. J., et al. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *PNAS*, 115, 6171–6176. [doi:10.1073/pnas.1721355115](https://doi.org/10.1073/pnas.1721355115)
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16, 252–257. [doi:10.3758/PBR.16.2.252](https://doi.org/10.3758/PBR.16.2.252)
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *CVPR*. [arXiv:1503.03832](https://arxiv.org/abs/1503.03832)
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626. [doi:10.1126/science.1110589](https://doi.org/10.1126/science.1110589)
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A*, 43, 161–204. [doi:10.1080/14640749108400966](https://doi.org/10.1080/14640749108400966)

DATA, LAW AND METHODS RESOURCES

- Center for Open Science. Preregistration. cos.io/initiatives/prereg
- General Data Protection Regulation, Article 9: Processing of special categories of personal data. gdpr-info.eu/art-9-gdpr
- National Institute of Standards and Technology. Face Recognition Vendor Test (FRVT). nist.gov
- UK Biobank. ukbiobank.ac.uk
- University of Adelaide (2015). One in a trillion odds of having a real-life doppelganger. adelaide.edu.au