

Characteristics Space

Foundations • The Valuation Engineer

Bert Craytor • July 9, 2026

Version 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

Formal definition. The characteristics space of a class of heterogeneous goods is the vector space whose coordinates index the attributes that buyers value, and in which each good is represented by the vector of its attribute levels.

Intuitive framing. The previous entry established that a heterogeneous good is a distinct bundle of characteristics. The natural next question is: how should we describe and compare such bundles formally? The answer, due to Lancaster (1966), is to treat each good as a point in a vector space whose axes are the characteristics buyers care about.

In this representation, a Pacifica residential property is no longer “a house in Park Pacifica” but a vector with coordinates for gross living area, lot size, view, condition, school district, distance to the ocean, and so on. Two properties that share all of these attribute values would occupy the same point in characteristics space; properties that differ in any attribute occupy different points. The geometry of the space encodes how close or how different two properties are, in a way that captures all the dimensions on which buyers make distinctions.

The dimensionality of the space depends on what attributes are relevant to the market under study. For a typical residential appraisal, the relevant space might have a dozen dimensions: physical (GLA, lot, beds, baths, age, quality, condition), locational (school district, view, proximity to amenities or disamenities), and legal/economic (zoning, easements). For a commercial appraisal, the relevant dimensions shift toward cash-flow attributes (lease structure, tenant quality, occupancy), but the underlying geometric idea is the same.

What makes the characteristics space useful is that it converts a qualitative practitioner’s intuition (“these comps are similar to the subject”) into a quantitative claim (“these comps are close to the subject in characteristics space”). Once distance is defined, every sales comparison decision can be examined as a statement about location in the space.

Where appraisers encounter it. Appraisers work in characteristics space constantly, usually without naming it. Every line of a sales comparison grid is a coordinate axis; every comp is a row vector of coordinate values; every adjustment is an attempt to move along one axis while holding the others fixed.

The eight Pacifica comparables we will use throughout this issue illustrate the construction directly. Each comp is a point in a four-dimensional characteristics space spanned by GLA, lot size, view, and condition:

Comp	GLA (sf)	Lot (sf)	View	Cond.	Sale price
A	1,650	9,000	yes	good	\$1,425,000
B	1,650	7,200	no	good	\$1,280,000
C	1,850	7,500	no	good	\$1,407,500
D	1,450	7,000	no	good	\$1,155,000
E	1,700	8,000	yes	excellent	\$1,470,000
F	1,550	7,100	no	average	\$1,177,500
G	1,700	10,500	no	good	\$1,392,500
H	1,600	7,300	no	good	\$1,312,500

Comps A and B from the previous entry are the first two rows; the remaining six expand the example in ways we will exploit through the rest of the issue.

Each comp is a point in $\mathbb{R}^4$:

- Comp A is at (1650, 9000, 1, 2).
- Comp B is at (1650, 7200, 0, 2).
- Comp G is at (1700, 10500, 0, 2).

The view and condition coordinates are encoded numerically (view as $\{0, 1\}$, condition as $\{1, 2, 3\}$); the price is *not* a coordinate of the characteristics space but a function of position within it.

Why it matters for defensibility. The characteristics space framework sharpens what comp selection actually is. A comparable is not just “a similar property” but a point in the relevant region of characteristics space close to the subject. The closer a comp is to the subject in the space, the smaller the adjustments required and the more reliable the indicated value becomes.

This reframing exposes several practice questions that the informal grid obscures:

- *Which dimensions matter?* The choice of axes is a modeling decision. Including “view” as an axis only makes sense if the market prices view distinctly; including “Feng Shui orientation” only makes sense if a market segment prices it. Defensible appraisal requires defending the axes, not just the values along them.
- *What distance metric applies?* Two comps may differ by \$200,000 in price because they are far apart in characteristics space, but the distance computation requires a scaling for each axis. Untransformed Euclidean distance treats one square foot of GLA as equivalent to one square foot of lot, which the market does not.
- *Are the comps in a region where the price function is well-behaved?* If the subject sits in a sparsely-sampled corner of characteristics space, no comp lies close to it and the value indication is necessarily extrapolative.

Worked appraisal example. The eight Pacifica comps occupy a relatively compact region of the four-dimensional characteristics space. Consider the question of which comp is “closest” to a hypothetical subject property at (1,700, 7,500, 1, 2) — 1,700 sf, 7,500 sf lot, view, good condition.

If we naively use Euclidean distance with unscaled axes, the lot dimension dominates because its values (7,000–10,500) are numerically much larger than view (0–1) or condition (1–3). Distances to the subject:

Comp	Distance (unscaled Euclidean)	Notes
A	1,501	view matches; lot dominates
B	308	view differs; lot close
C	230	view differs; closest by raw distance
D	559	view differs; GLA off
E	502	view matches; condition differs
F	422	view and condition differ
G	3,000	view differs; lot far away
H	224	view differs; closest of all

By raw distance, Comp H is the nearest neighbor. But Comp H has no view and the subject does. The naive metric has misled us: the lot dimension’s numerical scale crowded out the categorical view dimension, which actually matters more to the price.

A defensible distance metric would scale each axis by its implicit price — roughly, by how many dollars one unit of variation in that axis is worth. Once the axes are scaled by implicit price, view (worth ~\$100,000) is no longer dominated by lot (~\$25 per square foot). Comp A, which matches the subject on view, then emerges as the appropriate primary comparable.

This is precisely what the implicit-price machinery developed in the following entries provides: a principled scaling of the characteristics space that turns “closeness” from a geometric guess into a market-derived computation.

References.

- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Cross-references: heterogeneous good; hedonic price function; implicit price.

Addendum: Prolog Formalization

The following machine-readable companion (002-characteristics-space.pl) formalizes this entry in the journal’s Prolog knowledge base. It is also published as a browsable galley alongside this article.

```
%% 002-characteristics-space.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 002
%% Title: Characteristics Space
%% Author: Bert Craytor
%% Concept: characteristics_space
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('.././../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
```

```

% term(characteristics_space, "Characteristics Space", economic_theory,
%      "The characteristics space of a class of heterogeneous goods is
%      the vector space whose coordinates index the attributes that
%      buyers value, and in which each good is represented by the
%      vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(heterogeneous_good, ...).
% term(hedonic_price_function, ...).
% term(implicit_price, ...).
%
% generalized_by(characteristics_space, lancaster_1966).
% depends_on(characteristics_space, heterogeneous_good).
% -----

:- use_terms([ characteristics_space, heterogeneous_good,
                hedonic_price_function, implicit_price,
                comp_selection, sales_comparison_approach,
                real_estate ]).
:- use_relations([related/2, depends_on/2, generalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(characteristics_space).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.002').

% =====
% LOCAL VOCABULARY
% =====
%
% Narrow article-local field names + the dimensions of the worked
% example's characteristics space.

local_term(dimension).
local_term(coordinate).
local_term(euclidean_distance).
local_term(distance_metric).
local_term(nearest_neighbor).

% =====
% FORMAL ATTRIBUTION
% =====

% The construction is due to Lancaster (1966), as noted in citations.
attributed_construction(characteristics_space, lancaster_1966).

% =====
% THE ARTICLE'S CLAIM ABOUT REPRESENTATION
% =====
%
% A heterogeneous good becomes a point in a vector space. Each axis is
% a characteristic buyers value; each good is the vector of its
% attribute levels.

representation_of(heterogeneous_good, point_in_vector_space).

% Examples of dimensions relevant to residential appraisal.

```

```

residential_dimension(gla_sqft).
residential_dimension(lot_sqft).
residential_dimension(view).
residential_dimension(cond).
residential_dimension(bedrooms).
residential_dimension(baths).
residential_dimension(age).
residential_dimension(quality).
residential_dimension(school_district).
residential_dimension(distance_to_amenity).
residential_dimension(zoning).

% Examples of dimensions relevant to commercial appraisal | included
% to make the article's point that the dimensions shift with market.
commercial_dimension(lease_structure).
commercial_dimension(tenant_quality).
commercial_dimension(occupancy).
commercial_dimension(cash_flow_profile).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% Mapping appraisal practice onto the characteristics-space framework.

practice_to_space_correspondence(
    sales_comparison_grid_line, coordinate_axis,
    "A grid line is an axis.").
practice_to_space_correspondence(
    comp_row, row_vector,
    "A comp is a vector of coordinate values.").
practice_to_space_correspondence(
    adjustment, movement_along_axis,
    "An adjustment moves along one axis holding others fixed.").

% =====
% WORKED EXAMPLE
% =====
%
% The 8 Pacifica comps are points in a 4D characteristics space spanned
% by GLA, lot_sqft, view, cond_numeric. Bring in the shared dataset.

:- consult('../..kb/pacifica_comps').

% Axes of the chosen space for this article's analysis.
axis(gla_sqft).
axis(lot_sqft).
axis(view).
axis(cond_numeric).

% Numeric encoding for non-numeric axes.
encode(view, no, 0).
encode(view, yes, 1).
encode(cond_numeric, 1, 1).
encode(cond_numeric, 2, 2).
encode(cond_numeric, 3, 3).

%! coord(?Property, ?Axis, ?Numeric) is nondet.
%
```

```

% Numeric coordinate of Property along Axis. Handles symbolic attribute
% values by encoding them.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V ; encode(A, V, N) ).

% point(?Property, -Vector) is det.
%
% Vector representation of Property in the 4D characteristics space.

point(P, [GLA, Lot, View, Cond]) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond).

% =====
% THE HYPOTHETICAL SUBJECT
% =====
%
% The article tests distance metrics against this hypothetical property.

hypothetical_subject(subject).
attribute_value(subject, gla_sqft, 1700).
attribute_value(subject, lot_sqft, 7500).
attribute_value(subject, view, yes).
attribute_value(subject, cond_numeric, 2).

% =====
% NAIVE EUCLIDEAN DISTANCE (the article's straw distance metric)
% =====
%
% Computes the unscaled Euclidean distance between Property and the
% hypothetical subject. Used to demonstrate that unscaled distances
% mislead: the lot dimension dominates because its numeric range is
% much larger than view (0..1) or cond (1..3).

naive_distance_to_subject(P, D) :-
    point(P, V_P),
    point(subject, V_S),
    sum_of_squares(V_P, V_S, S),
    D is sqrt(S).

sum_of_squares([], [], 0).
sum_of_squares([X|Xs], [Y|Ys], S) :-
    sum_of_squares(Xs, Ys, S0),
    D is X - Y,
    S is S0 + D * D.

% =====
% THE ARTICLE'S CENTRAL OBSERVATION
% =====
%
% Under naive Euclidean distance, Comp H is the nearest neighbor. But
% Comp H has no view and the subject does; view matters more to price
% than the naive metric admits. The fix (developed in entries 003-005)
% is to scale each axis by its implicit price.

```

```

claim(naive_metric_misleads,
    "Unscaled Euclidean distance crowds out small-range axes (like view) under large-range axes (
    ↪ like lot_sqft). A market-defensible distance scales each axis by its implicit price.>").

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====
%
% Three questions the characteristics-space framing makes explicit.

defensibility_question(which_dimensions,
    "The choice of axes is a modeling decision. Defending an appraisal requires defending the axes,
    ↪ not just the values on them.").
defensibility_question(distance_metric,
    "Untransformed Euclidean distance treats one sqft of GLA as equivalent to one sqft of lot,
    ↪ which the market does not. Distance requires a scaling per axis.>").
defensibility_question(extrapolation_region,
    "If the subject sits in a sparsely-sampled corner of the space, no comp is truly close and the
    ↪ indication is extrapolative.>").

% =====
% INFERENCE RULES
% =====

%! same_point(?A, ?B) is nondet.
%
% Two properties occupy the same point in the characteristics space.
% (Doesn't occur in the running example, but the article notes that it
% would if all coordinates matched.)

same_point(A, B) :-
    point(A, V),
    point(B, V),
    A \== B.

%! nearest_neighbor_by_naive_distance(-P) is semidet.
%
% The comp closest to the subject under the naive Euclidean metric.
% Used to illustrate the article's point that the naive metric chooses
% Comp H, not Comp A.

nearest_neighbor_by_naive_distance(P) :-
    findall(D-Q,
        ( property(Q), Q \== subject, naive_distance_to_subject(Q, D) ),
        Pairs),
    keysort(Pairs, [_P|_]).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```