CrossMark

Do design decisions depend on "dictators"?

David A. Broniatowski[1]

Abstract Design decisions often require input from multiple stakeholders or require balancing multiple design requirements. However, leading axiomatic approaches to decision-based design suggest that combining preferences across these elements is virtually guaranteed to result in irrational outcomes. This has led some to conclude that a single "dictator" is required to make design decisions. In contrast, proponents of heuristic approaches observe that aggregate decisions are frequently made in practice, and argue that this widespread usage justifies the value of these heuristics to the engineering design community. This paper demonstrates that these approaches need not be mutually exclusive. Axiomatic approaches can be informed by empirically motivated restrictions on the way that individuals can order their preferences. These restrictions are represented using "anigrafs"—structured relationships between alternatives that are represented using a graph–theoretic formalism. This formalism allows for a computational assessment of the likelihood of irrational outcomes. Simulation results show that even minimal amounts of structure can vastly reduce the likelihood of irrational outcomes at the level of the group, and that slightly stronger restrictions yield probabilities of irrational preferences that never exceed 5%. Next, an empirical case study demonstrates how anigrafs may be extracted from survey data, and a model selection technique is introduced to examine the goodness-of-fit of these anigrafs to preference data. Taken together, these results show how axiomatic consistency can be combined with empirical correspondence to determine the circumstances under which "dictators" are necessary in design decisions.

Keywords Anigraf · Arrow's theorem · Mental models · Decision-based design · Simulation · Trajectory mapping

✉ David A. Broniatowski
  broniatowski@gwu.edu

[1] Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, 800 22nd St. NW #2700, Washington, DC 20052, USA

1 Motivation

Design often requires balancing the competing needs of multiple stakeholders or design requirements. Stakeholders may not agree on the characteristics of the best design; that is, their preferences may differ. Similarly, a design which aims to best fulfill one requirement may not perform adequately on other requirements. Thus, the "preferred" design may differ across these elements.

A considerable body of work has examined the impact of differing preferences on design. On one hand, Hazelrigg (1996, 1997, 1998, 1999) argues that defining a "best" rational design (i.e., one that maximize some utility function) requires attending only to a single decision-maker (a "dictator"). This position is justified using mathematical theories of social choice. Given a set of generally accepted axioms, Arrow's Impossibility Theorem (Arrow 1963; described below) states that one cannot guarantee the existence of a rational aggregate preference ordering (and therefore, an aggregate utility function) when there are at

Springer

least three alternatives. Specifically, attempts to aggregate preferences across individuals may result in *cyclic preferences* at the level of the group, such that one alternative is both superior and inferior to another. Furthermore, even if there is such a "dictator", Franssen (2005) argues that most decision-based design methods are subject to failure when that individual must choose between concepts with multiple competing requirements because these design requirements function exactly as the stakeholders do in Hazelrigg's version of the problem.

On the other hand, scholars such as Frey et al. (2009) argue that these axiomatic arguments are not empirically justified. They point to the wide use of popular techniques such as the Pugh Controlled Convergence (PuCC) method (Pugh 1991), the Analytic Hierarchy Process (AHP; Saaty 1990), and similar techniques, arguing that engineers' satisfaction with the decision outcome is a better measure of a technique's utility than is any mathematical guarantee of optimality or rationality.

A recent editorial in *Research in Engineering Design* (Reich 2010) characterized this debate as one between "scientism" and "praxis" (or "coherence" and "correspondence"; Katsikopoulos 2009, 2012). Specifically, axiomatic approaches aim to maintain internal consistency such that the *process* used to make design decisions is logically consistent. In contrast, practical approaches aim to maximize external validity, such that the *results* of the design process are widely accepted by the design community, regardless of whether they might be chosen in a rational manner (e.g., by allowing a group utility function to be maximized or even defined). Thus, the core of the debate centers around differences in how a design should be evaluated.

Advocates for the axiomatic approach argue that all design decisions must be made by a single decision-maker, and that alternate methods are virtually guaranteed not to deliver value to the customer in the long run. Furthermore, these alternatives leave designers subject to the caprices of known social processes such as fads, groupthink, and status effects; especially when experts disagree. Finally, even if there is a single decision-maker, Franssen (2005) argues that any multi-criteria decision problem is still subject to irrational outcomes because the criteria themselves may be treated as independent decision-makers. In contrast, proponents of the heuristic approach argue that techniques lacking axiomatic coherence, such as PuCC and AHP, have been shown to perform well in practice despite theoretical limitations.

This paper draws upon recent innovations in the social choice and mathematical psychology literatures, and especially the work of Richards et al. (1998, 2002) to demonstrate how these approaches might be synthesized. This demonstration relies upon a slight, yet empirically

supported, relaxation of one of Arrow's axioms: the axiom of unrestricted domain. Given this relaxation, one may define a set of conditions under which a group of decision-makers can rationally aggregate their preferences. Specifically, if teams of designers are known to share a common mental model, one may often define an aggregate set of preferences for the group. This group-level preference ordering exists under the vast majority of circumstances, allowing team members to select a "best" design from among a set of structured alternatives. A similar argument applies to comparisons between design requirements that are related by known physical laws and other sources of structure specific to the design domain.

The novelty of the approach presented here comes from the use of designers' mental models (typically associated with empirical regularities) that constrain the set of possible preference orderings between alternatives considered by designers. Although this constraint violates one of the axioms typically taken for granted in decision-based design, such a violation need not result in an irrational design. On the contrary, this approach rules out "irrational" preference orders that violate empirical regularities and are therefore "unthinkable". Such an approach allows for the incorporation of empirical regularities (including physical constraints, cognitive effects, and socio-cultural factors) into the decision-making process, such that decision-makers choose between alternatives based on the knowledge they have available to them. Importantly, this approach does not require that only one decision-maker's preferences should be followed to the exclusion of all others (i.e., the requirement for a "dictator"). Rather, decision-makers must only agree on the underlying problem structure—a much weaker condition. This slight relaxation of the axiomatic approach to decision-making is shown to be sufficient to guarantee the existence and selection of an optimal[1] decision outcome in the vast majority of cases. Furthermore, this approach allows the precise definition of the conditions under which cyclic preferences (and therefore, no best outcome) may occur. Finally, this approach is used to motivate a new research agenda: the empirical measurement and evaluation of designers' mental models, how closely these mental models correspond to one another and to empirical regularities, and the implications of this correspondence for decision-based design.

The outline of this paper is as follows: Sect. 2 provides a motivation for the approach presented in this paper by comparing and contrasting leading approaches to decision-

---

[1] Here, "optimal" means that an aggregate preference order can be defined, that a top choice can be defined for this preference order, and that the method used to generate this preference order is most likely to result in a preference ordering that is consistent with the group's common design goal, as defined by Condorcet's formulation discussed in Young (1995). Importantly, Condorcet notes that the existence of such an optimum is not guaranteed.

based design. Section 3 introduces the main approach used in this paper. Section 4 presents the results of a simulation illustrating the generality of this approach. Section 5 demonstrates an application of this approach based on empirical data. Finally, Sect. 6 discusses this approach in light of alternate theories of decision-based design, and concludes by summarizing the novel contributions of this work.

2 Background

Engineers rely heavily upon techniques that help designers to select one or a subset of best designs from within a much larger tradespace. Since design teams may have many members, and the designs themselves may aim to meet several requirements or to incorporate feedback from several stakeholders, commonly used techniques aim to aggregate preferences across these elements. Examples of such techniques include the Analytic Hierarchy Process (AHP; Saaty 1990), which prompts designers to assign weights to decision criteria and to various stakeholders, and the Pugh Controlled Convergence Method (PuCC; Pugh 1991), in which teams of design engineers must reach consensus regarding pairwise comparisons between candidate designs and a baseline "datum" concept. Each such approach relies upon the aggregation of preferences from individuals to generate a group-level preference ordering for the design team.

2.1 Decision-based design

These methods have been sharply criticized by adherents of the decision-based design paradigm, and especially Hazelrigg (1996), who claims that "…popular approaches to design optimization, such as Total Quality Management (TQM) and Quality Functional Deployment (QFD), are logically inconsistent and can lead to highly erroneous results." (p. 161). To illustrate this point, Hazelrigg considers the case where a team of decision-makers must choose between three alternatives: A (e.g., an apple), B (a pear), and C (an orange). For an individual who prefers A to B, B to C, and A to C, we say that A > B > C. Furthermore, one would expect this decision-maker to choose option A all else being equal. Following Arrow (1963), Hazelrigg next considers the hypothetical scenario where there are three decision-makers. The first decision-maker has preference order A > B > C, the second B > C > A, and the third C > A > B. Under these conditions, and given a choice between A and B, two decision-makers would "vote" for option A such that the preference of the group is A > B. Similarly, the group prefers B > C, and C > A. Combining these elements, we see that A > B > C > A—a "cyclic" preference structure. When

cyclic preferences such as these exist, there is no best design. Furthermore, Hazelrigg (1996) asserts that selection of any design that is part of such a preference cycle would be disastrous:

> Given an apple, the individual [with cyclic preferences] would prefer to trade to a pear. And since the individual has a strong preference for a pear, he would be willing to give up something of value to affect the trade, say a penny. After the trade, the individual has a pear. But he would prefer to have an orange, and would be willing to spend another penny to affect this trade. Now, with the orange, he would spend another penny to trade for an apple, returning to his starting point at a loss of three cents. This trading would continue ad infinitum…He cannot stop trading without denying his preferences …the case shown here is not a rare, pathological case. It is the norm…Virtually all groups will have intransitive preferences…As a consequence, any methodology that demands the construction of a group utility function in any aspect of its construct is logically inconsistent and doomed to failure (p. 162).

This critique is based on Arrow's Impossibility Theorem (1963), a seminal result in social choice theory. Arrow's Impossibility Theorem states that there is no ranked voting system that can convert individual preferences to a complete and transitive ranking for the group (i.e., one in which all preferences are ranked and there are no cyclic preferences) that also obeys the following axioms:

- Unrestricted domain: All individual preference orders are admissible. For example, one cannot say that B > C > A is "not allowed". All individual rankings are allowed.
- Independence of irrelevant alternatives: The social order will not change by addition of a new alternative. For example, if the group prefers A > B when this pair is considered in isolation, the group will still prefer A > B when A is compared to C and when B is compared to C.
- Monotonicity: If an individual changes his or her preference by ranking a given alternative more highly, that alternative cannot be ranked less highly in the social order as a result. For example, if the social order is A > B > C, and an individual changes his or her preference from C > B > A to B > C > A (i.e., C is now ranked more highly), the rank of C cannot go down in the social order (i.e., A > C > B is not an admissible social order).
- Non-imposition: Any function that generates a social order from individual orders should be able to generate all
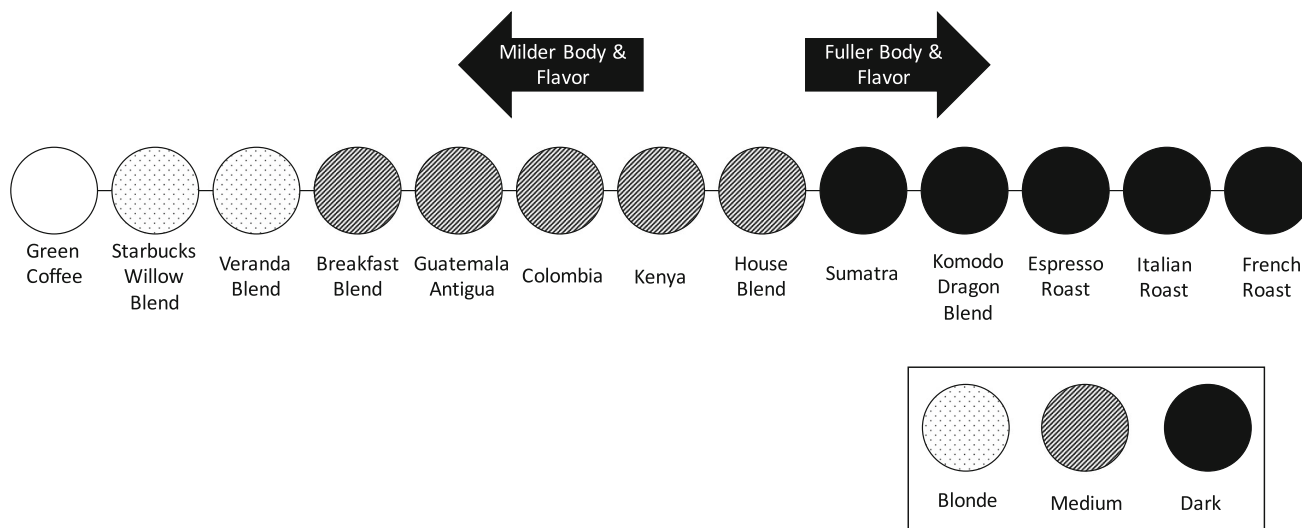
Fig. 1 The Starbucks "roast curve"—a set of alternatives that are structured so as to induce single-peaked preferences

social orders for some set of individual orders. For example, given alternatives A, B, C, and D, there must be some set of individual orders such that A > B > C > D.

- Non-dictatorship: There is no individual for whom the social order always mirrors that individual's preference order.

As the number of decision-makers and alternatives grows, the likelihood of encountering cyclic preferences converges to unity. Since individual decision-makers are assumed not to have cyclic preferences, Hazelrigg interprets Arrow's Theorem as a proof that a "dictator"—i.e., one decision-maker who determines the final preference order—is necessary for a rational design. Strictly speaking, Arrow's Theorem does not specify that this individual must be a literal dictator. Rather, it is a formal requirement of the axiomatic framework that there is a one-to-one correspondence between this individual's preferences and the group's preferences, regardless of aggregation method, and presuming all the other axioms are correct. Nevertheless, Hazelrigg (1997) states:

> It is the responsibility of the project manager to align, to the maximum possible extent, the utilities of the individual designers. This can be done in two ways. First, it is necessary for the manager to state explicitly the utility against which the design should be judged. Second, the manager must create a set of incentives and rewards that make it in the best interest of each design engineer to make use of the stated utility measure (p. 196).

2.1.1 The axiom of unrestricted domain

Arrow's axiom of unrestricted domain specifies that any limitation placed on the set of possible preference

orderings could potentially negate the impossibility result. Significant attention in the literature has focused on a specific class of domain restrictions: "single-peaked preferences" (Black 1948). Single-peaked preferences occur when alternatives may be ranked relative to one another in a linear fashion. For example, jet engines may be ranked relative to one another in terms of their specific impulse. Similarly, Starbucks uses its "roast curve" to rank coffee flavors along a left-to-right spectrum (see Fig. 1). Building upon this insight, Scott and Antonsson (1999) have argued that the axiom of unrestricted domain is not applicable to engineering design because any given engineering variable is almost always ordered on some external scale such that "less is better, more is better, or closer to a particular target is better" (p. 224). Per the coffee example, an individual decision-maker who prefers a Sumatra blend would choose a House Blend over a Veranda Blend when the Sumatra blend is unavailable. Similarly, a vehicle structures group may prefer a material with a stronger bending stiffness over a weaker one. Preference orders that are inconsistent with this structure, such as a preference for a French Roast over an Italian Roast, but green coffee over French Roast, would be considered irrational from an empirical perspective. Thus, Scott and Antonsson contend that Arrow's Theorem does not apply to decision-based design, and that a choice that is best for the group should exist if each individual criterion is single-peaked. There is some theoretical support for this position. For example, Barberà, Gul, and Stacchetti (1993) have generalized single-peaked preferences to multiple dimensions, showing that Arrow's Theorem does not apply when preferences may be ranked relative to one another in a grid, such as when designs can be located relative to one another in a multidimensional tradespace with a global optimum. Indeed, Scott and

Antonsson note that designers tend to avoid unrestricted domain problems by restricting their attention to local optima (treating them like global optima).

2.1.2 Arrow's theorem in the context of multi-criteria decision-making

Contrary to Scott and Antonsson (1999), Franssen (2005) argues that single-peaked preferences at the level of design criteria are insufficient to avoid cyclic preferences. This is because they do not, in general, imply single-peaked preferences at the level of the design task. On one hand, he notes that conceptual design requirements fulfill the same function as human decision-makers in social choice problems; on the other hand, he asserts that these requirements behave in a manner that is fundamentally unlike humans. Specifically, Franssen claims that each criterion (regardless of whether it is single-peaked) can order its preferences differently per its own internal logic. If these criteria are treated as independent "decision-makers" they are subject to Arrow's paradox. Indeed, it is traditional to think of design criteria as orthogonal (i.e., independent) axes in a tradespace, where orthogonality implies that the associated criteria are uncorrelated.[2]

It is well known that design performance criteria are often correlated to some degree. Consider a system with requirements for response time and accuracy. The well-known speed/accuracy tradeoff in control theory states that any system that aims to increase its response time must sacrifice its ability to avoid oscillation or overshoot, all else being equal. This relationship between these two design criteria violates Franssen's assumption of independence between design requirements. Similarly, a system's maintainability (as measured by lines of code or cyclomatic complexity) generally comes at the cost of robustness or flexibility—the ability of the system to respond to unanticipated changes (see Broniatowski and Moses 2016 for a more general discussion of the relationship between flexibility and system complexity). Finally, consider the design of an airplane, where one must typically make fundamental tradeoffs between criteria such as cruising speed and mass. These tradeoffs are often direct consequences of the laws of physics and, even though these designs are typically represented by coordinates in a high-dimensional tradespace, the designs themselves are not uniformly distributed throughout that space, as would be expected if these criteria were indeed independent. Clearly, not all design

criteria are correlated with one another. However, such correlations, when they exist, place limits on the ability of these criteria to order their "preferences" across designs. Indeed, these physical relationships between design criteria are at least as constraining as the social constraints on preference orders.

Franssen observes that a system's physical form is in many cases distinct from its use, or function (see also Broniatowski 2017), and that preferences over design criteria are driven by a decision-maker's mental representation of these criteria more so than their physical attributes. This is perhaps most clearly expressed in his statement that "preference is a mental concept and is neither logically nor causally determined by the physical characteristics of a design option" (Franssen 2005, pp. 48–49). Furthermore, he warns against imposing artificial constraints upon the set of admissible preference orders because they can "… be interpreted as compromising the impartiality of the procedure, when it reflects a restriction of the freedom of the individuals to order the options as they please." (p. 44). Although physical constraints limit the design of the form, the underlying assumption is that mental concepts are not constrained except by artificial means, and therefore, all preference orders are admissible, consistent with the axiom of unrestricted domain—a chain of logic that leads inexorably to Arrow's Impossibility result.

2.2 Mental models

Mental concepts are indeed constrained. Furthermore, these constraints are often natural, not artificial. Several bodies of scholarly literature have independently concluded that structured mental constraints, or "mental models", are often held in common by members of a design team (e.g., Ahamed et al. 2016; Anderson 1995; Avnet and Weigel 2013; Langan-Fox et al. 2000, 2004). Structured mental models are a result of deep domain expertise that are strongly shaped by empirical regularities (e.g., Bang et al. 2007; Reyna and Lloyd 2006; Romney et al. 1996, 1986). For example, mental models may result from causal regularities in a system's behavior (Moray 1990), from sources of non-random structure in the environment, or from other repeated lawful behaviors (i.e., "natural modes", Richards 2008; Richards and Bobick 1988), including from the physical constraints mentioned above. Mental models are therefore highly structured and generally take the form of causal or similarity relations among categories of objects (Richards 2001). Richards et al. (1998) note that:

> Such knowledge structures and relationships are present in all choice contexts, whether physical, social, or cognitive. Their origin lies in the laws and regularities that

---

[2] If so, then Scott and Antonsson's (1999) argument, that designers tend to avoid unrestricted domain problems by restricting their attention to local optima, seems to somewhat refute Franssen's (2005) argument for many real designs. However, one could certainly conceive of designs where there are several local optima – the situation to which Franssen presumably refers.
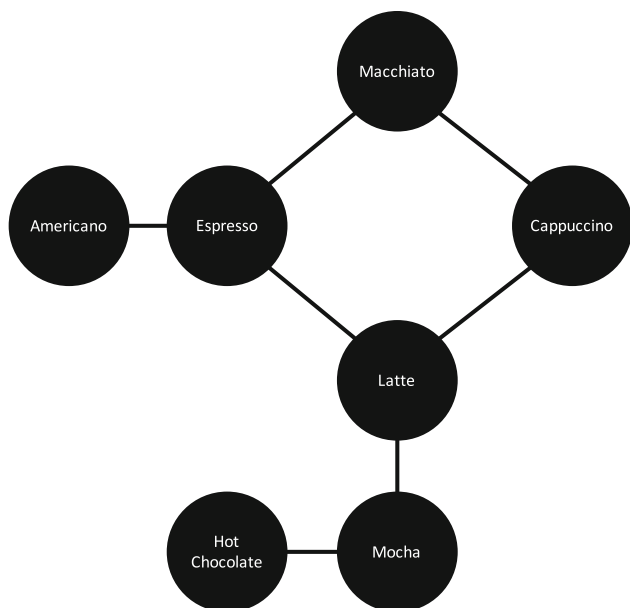
Fig. 2 An anigraf representing similarity relationships between different types of hot beverages

bring order into our world…To illustrate, consider the frequency of vocal sounds made by animals: large animals make low pitch sounds because they have large vocal tracts, whereas small animals with smaller vocal cavities will make higher pitched sounds. If we hear a sound made by an unseen animal in the forest, we can guess the size, and hence the category of the animal. We all share and use this kind of intrinsic knowledge in order to make rational perceptual inferences from sense data. Second, at a more cognitive level, stories, like other forms of linguistic communication, must have a known intrinsic structure in order for the meaning to be understood by the audience…consider types of stories as reflected in categories of films. Film categories have a natural ordering from "light cognitive" such as romantic comedies to "heavy physical" such as martial arts. People who prefer light cognitive films will typical avoid violent films, and vice versa. However, both groups of people may accept documentaries…Finally, our daily social interactions are also very regularized and lawful, following certain traditions and conventions. In the U.S., we drive on the right side of the road, with the steering wheel on the left. In Japan and Britain, it is the opposite. These conventions dictate the placement of traffic signals, signs, and which way we look first before crossing the street. Such strong correlations at all levels—perceptual, cognitive, and social—impose an enormous amount of structure on our thoughts and behaviors, and affect our ways of holding and sharing knowledge (p. 3).

2.2.1 Anigrafs

Mental models structure relationships between alternatives. In this section, this paper shows how the structure of these relationships can impose restrictions on the set of possible preference orders that may be entertained, thereby avoiding cyclic preferences. As an illustration, consider seven drinks which one may obtain from a coffee shop (see Fig. 2):

- Espresso: Dark-roasted filtered coffee
- Americano: An espresso mixed with hot water
- Macchiato: An espresso mixed with foamed milk
- Latte (or Flat White): An espresso mixed with steamed milk
- Cappuccino: Equal parts espresso, foamed milk, and steamed milk
- Mocha: An espresso mixed with steamed milk and chocolate syrup
- Hot chocolate: Steamed milk mixed with chocolate syrup

Given this illustrative example, a latte is more similar to a mocha (a mocha is simply a latte with chocolate syrup) than it is to an americano. Furthermore, given a choice between these seven alternatives, and all else being equal, a consumer with a preference for a mocha might rank a latte as his or her second choice, e.g., if chocolate syrup has run out. In contrast, an espresso would be ranked lower and an americano even lower.

The same logic holds in the domain of multi-criteria decision-making. Even if one were to treat design criteria as fully independent "decision-makers", the preference orderings over design options corresponding to each individual criterion constrain one another to the extent that the criteria are themselves correlated, as in the relationships between speed and accuracy discussed above (in contrast, the speed of a system is usually not related to its color). Because of these correlations, it would be irrational (in the sense that it violates the laws of physics) to prefer more speed without also preferring less accuracy (and vice versa).

Figure 2 is an example of what Richards (2015) calls an "anigraf"—a graph-based representation of a mental model encoding similarity relations between alternatives. In general, given that an individual expresses a preference for a top-ranked alternative, the anigraf imposes a partial order on the remainder of that individual's preferences. Table 1 shows the preference ordering expected for each item in the graph. For example, an individual that prefers a hot chocolate would rank a mocha as second best, a latte as third best, etc.

Table 1 Admissible individual-level preference orderings for agents given the anigraf in Fig. 2

1st choice	2nd choice	3rd choice	4th choice	5th choice
Americano	Espresso	Macchiato or Latte	Cappuccino or Mocha	Hot chocolate
Espresso	Americano, Latte, or Macchiato	Cappuccino or Mocha	Hot chocolate	
Macchiato	Espresso or Cappuccino	Latte or Americano	Mocha	Hot chocolate
Cappuccino	Macchiato or Latte	Espresso or Mocha	Americano or hot chocolate	
Latte	Espresso, Cappuccino, or Mocha	Hot chocolate, Americano, or Macchiato		
Mocha	Latte or hot chocolate	Espresso or Cappuccino	Macchiato or Americano	
Hot chocolate	Mocha	Latte	Espresso or Cappuccino	Americano or Macchiato

Table 2 Weights to each vertex in Fig. 2 anigraf

Americano (A)	Espresso (B)	Macchiato (C)	Cappuccino (D)	Latte (E)	Mocha (F)	Hot chocolate (G)
0	5	3	4	4	0	9

The weight of each vertex is equivalent to the number of respondents for whom the corresponding option is their most preferred choice

2.2.2 Condorcet tallies

Anigrafs represent the structure imposed upon a set of alternatives, restricting the domain of preference orderings. Furthermore, Richards (2015) describes how one may use an anigraf to determine an aggregate partial order over preferences for a group. For example, consider a hypothetical group with 25 members that must select one type of beverage for all members. The number of group members preferring each flavor is shown in Table 2. Given these preferences, the optimal aggregate group preference may be determined by means of a summation over pairwise Condorcet tallies (de Caritat marquis de Condorcet 1785), corresponding to Saari's "pairwise plurality rule" (Saari 1994; Saari and Sieberg 2004)—i.e., the winner is the option that is preferred in the largest number of pairwise comparison, as shown in Table 3.

Although Arrow showed that such tallies are, in general, subject to cyclic preferences when the domain is unrestricted, domain restrictions can alleviate this problem. For example, this particular tally yields a total aggregate preference order for this group with no cycles: E > F > B > D > C > G > A. Although only 4 out of 25 (16%) group members identified a latte as their most preferred alternative, it is the highest-ranked in the aggregate preference order and therefore represents the optimal choice. Specifically, when the group chooses a latte, four people (16%) get their top choice, thirteen people (52%) get their second choice or better, and all 25 people (100%) get their third choice or better. Furthermore, the group's second most preferred choice is a mocha—a compromise between a hot chocolate and a latte—even though no group members chose the mocha as the best alternative and hot chocolate, the option preferred by a plurality of nine group members (36%), is ranked second worst overall.

2.2.3 Indifference between options

When group members must select between multiple options that are equidistant from their top choice, the anigraf method predicts that those members will, on balance, be indifferent between these options. Although individual group members might express "micro-preferences"—e.g., some of those possessing an espresso as their first choice would prefer an americano over a latte when these are directly compared, whereas others would have the opposite preference—recall that the overall preference order for the group is constructed using a Condorcet tally: one determines the overall group preference by subtracting the number of people who prefer the latte from the number of people who prefer the americano. Thus, the anigraf model would predict that, at the level of the group, members are indifferent between these options. In practice, there might be slight differences between predictions and actual subject data due to measurement error. If one assumes that this error is symmetrically distributed, meaning that the number of people preferring americano as their second choice is roughly equal to the number of people choosing latte as their second choice, these two groups cancel each other out in the Condorcet tally, leading to zero contribution overall.[3]

---

[3] One might object to the assumption of a symmetric distribution over these two options. However, if large and consistent preferences for one option over the other exist, the anigraf itself would possess a different structure. Thus, in such a case, we are no longer dealing with "micro"-preferences at all.

Table 3 Condorcet tally results given weights shown in Table 2

	A (0)	B (5)	C (3)	D (4)	E (4)	F (0)	G (9)	Total	Preferred outcome
A vs B	0	−5	−3	−4	−4	0	−9	−25	B
A vs C	0	0	−3	−4	0	0	0	−7	C
A vs D	0	5	−3	−4	−4	0	−9	−15	D
A vs E	0	0	0	−4	−4	0	−9	−17	E
A vs F	0	5	3	−4	−4	0	−9	−9	F
A vs G	0	5	3	0	0	0	−9	−1	G
B vs C	0	5	−3	−4	4	0	9	11	B
B vs D	0	5	0	−4	0	0	0	1	B
B vs E	0	5	3	−4	−4	0	−9	−9	E
B vs F	0	5	3	0	0	0	−9	−1	F
B vs G	0	5	3	4	4	0	−9	7	B
C vs D	0	5	3	−4	−4	0	−9	−9	D
C vs E	0	0	3	0	−4	0	−9	−10	E
C vs F	0	5	3	4	−4	0	−9	−1	F
C vs G	0	5	3	4	0	0	−9	3	C
D vs E	0	−5	3	4	−4	0	−9	−11	E
D vs F	0	0	3	4	0	0	−9	−2	F
D vs G	0	5	3	4	4	0	−9	7	D
E vs F	0	5	3	4	4	0	−9	7	E
E vs G	0	5	3	4	4	0	−9	7	E
F vs G	0	5	3	4	4	0	−9	7	F

Each hot drink is followed by its corresponding weight, shown in parentheses as given in Table 2. Someone with a top choice for a given drink, X, will prefer closer drinks over farther drinks on the anigraf. Therefore, in a pairwise comparison between two drinks, the closer drink gets an equal number of "votes" to the total weight of X. Positive entries indicate that the first option in the pairwise comparison is preferred, whereas negative entries indicate that the second option is preferred. In the case where both drinks are equidistant from X, neither option gets any weight because, on average, the votes in favor of each option, if any, cancel each other out

A Americano, *B* Espresso; *C* Macchiato, *D* Cappuccino, *E* Latte, *F* Mocha, *G* Hot chocolate

2.2.4 Top cycles

Not all anigrafs preclude cyclic preferences. For example, if two more people preferring a cappuccino join the group, a rather complex cycle emerges: (D ∼ F) > C > B > (D ∼ F), meaning that the group is indifferent between a cappuccino and a mocha, prefers both of these to a macchiato, prefers a macchiato over an espresso, and prefers an espresso over both a cappuccino and a mocha.[4] However, cycles need not be a source of concern if it is still possible to define a single most preferred outcome. Despite the existence of a cycle, the group can identify option E, a latte, as the best choice overall, meaning that the group is still able to make a stable selection.

Richards et al. (1998, 2002) focused their attention on structures that induce *top cycles*: those for which it was not possible to define a single best alternative due to the existence of cyclic preferences among the top-ranked choices. For example, these authors showed that all rings with at least five nodes yield top cycles for some set of weights. This approach is extensible to engineering design teams: although one cannot generally control the preferences of a given team's members, one can determine empirically whether the team's preferences interact with the knowledge structure in a manner that can lead to top cycles.

3 Simulation

Richards et al. (2002) used a simulation to examine the circumstances under which randomly selected anigrafs might lead to top cycles. Their major finding was that even very limited domain restrictions greatly reduce the

---

[4] The pairwise plurality rule generates a partial order over options, and not always a total order. In this case, the partial order is E > {(D ∼ F) > B > C > (D ∼ F)} > G > A, where E is the top choice, A is the last choice, G is second-to-last, and the remaining options are all ranked worse than E but better than G and A. In addition, the group is indifferent between D and F. However, this is only a partial order, rather than a total order, because the relationships between F, B, C, and D lead to cycles, precluding an aggregate ranking of these alternatives relative to one another.

probability of top cycles. In this section, their results are replicated and extended.

The structure of the simulation proposed by Richards et al. (2002) is as follows: An Erdös-Rényi random graph (i.e., a graph in which each edge is present with fixed probability, p) is generated with n vertices. Weights, randomly selected between 0 and 1000, are assigned to each node at random following a uniform distribution. The result is an anigraf with n nodes. For each such anigraf, the Condorcet tally procedure is used to determine if there is a top cycle. Thus, by generating many such graphs, one can compute the *a priori* probability that a given graph with a given number of nodes will have a top cycle.

3.1 Knowledge depth

One might argue that it is unreasonable to expect engineering designers to possess fully ordered preferences over several design alternatives. This concern is especially relevant if there are many alternatives, if alternatives are closely preferred to one another, or if there is another source of noise in preference gathering data. Instead, one might expect designers to only have strong preferences over their top k alternatives. Following Richards et al. (2002), this paper refers to k as "knowledge depth." A knowledge depth of k means that decision-makers order the top $k + 1$ options per the anigraf, and all other options are ordered at random (following a uniform distribution). Thus, $k = 0$ means that the anigraf does not factor into decision-making at all and any pairwise comparisons between options that are not top-ranked are made uniformly at random, corresponding to an unrestricted domain. When $k = 1$, the top-ranked option is assigned a value of 0, the second best options (according to the anigraf) are assigned a value of 1, and all other options are assigned a random value between 2 and N. This means that each decision-maker's second best choice is the set of nodes that are one step removed from the top choice, and all other choices are determined uniformly at random. When $k = 2$, the top-ranked option is assigned a value of 0, the second- and third best options are assigned values of 1 and 2, respectively, and all other options are assigned a random value between 3 and N. Therefore, the second- and third best choices are determined by the anigraf and all other choices are determined uniformly at random, etc.[5] Thus, as

k increases, preferences are increasingly constrained by the anigraf.

In practice, one might speculate that k represents the maximum number of alternatives that a decision-maker can maintain in short-term memory. Although initially estimated as 7 ± 2 (Miller 1956), more recent estimates place human short-term memory capacity at 4 ± 1 (e.g., Mathy and Feldman 2012).[6] However, this speculative interpretation of k is not crucial to the results presented here.

3.2 Simulation procedure

Three parameters, n, p, and k, were used as inputs to the simulation described above. Specifically, 100,000 graphs were generated for values of n ranging from 3 to 20 in increments of 1, and 20 to 100 in increments of 10. Values of p ranged from 0.1 to 0.9 in increments of 0.1 ($p = 0.99$ was also included). Following Richards et al. (2002), the probability of top cycles was calculated for anigrafs in which agents had fully unconstrained preferences ($k = 0$), minimally constrained preferences ($k = 1$; $k = 2$ was also tested with results that did not differ significantly from $k = 1$), constraints consistent with human short-term memory limitations ($k = 3$), and preferences that were fully constrained by the associated anigraf. Results indicate that the probability of a top cycle was insensitive to changes in p, replicating the findings of Richards et al. (2002). Therefore, results were collapsed across this dimension, leading to 1 million samples for each value of k and n. Simulation results are shown in Fig. 3.

3.3 Simulation results

The analysis presented here suggests that cyclic preferences, although possible in theory, are a rare occurrence. Results show that, when preferences are unconstrained, the probability of a top cycle approaches 100%, especially for large numbers of nodes. However, even a minor constraint, $k = 1$, drastically decreases the probability of a top cycle. Even for 100 nodes, this probability remains below 5%. When $k = 3$, which might be interpreted as a lower bound for the value that is most consistent with human cognitive limitations, results are virtually indistinguishable from a fully constrained anigraf for a large number of nodes. Furthermore, for 100 nodes, the probability of a top cycle does not exceed 2%. Finally, the peak of the curve indexing

---

[5] Richards et al. (2002) distinguished between "knowledge depth", as defined here, and "knowledge depth with indifference", meaning that only the top k options are ranked, and all remaining options are given rank $k + 1$. The simulations in this manuscript all assume knowledge depth without indifference. Nevertheless, the results presented here do not differ qualitatively from those presented by Richards et al. (2002) who assumed knowledge depth with indifference.

[6] Experts may be able to hold many more alternatives in memory because their expertise allows them to "chunk" or abstract sequences of alternatives more efficiently (e.g., Mathy and Feldman 2012). In such a case, knowledge depth would be much greater than 3, which would impose even more structure on outcomes. As will be shown below, this means that experts are even more likely to avoid cyclic outcomes compared to novices.

Fig. 3 Simulation results represent one million runs per datapoint. The horizontal axis represents n, the number of vertices. The vertical axis, representing the probability of top cycles, uses a logarithmic scale
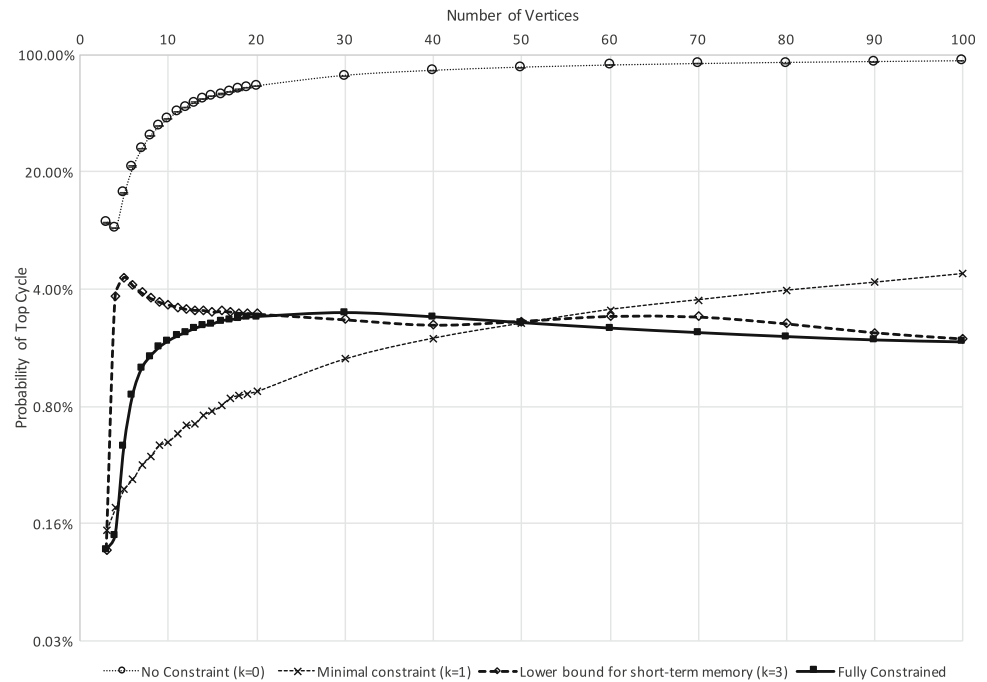
$k = 3$ never exceeds 5%. These results broadly replicate, and strengthen, the findings of Richards et al. (2002).

4 Obtaining anigraf data

Ultimately, anigrafs are mental constructs that must be elicited from groups of stakeholders. Richards and Koenderink (1995) proposed "Trajectory Mapping" (TM) as a non-metric scaling technique that may be used to elicit anigrafs from one or several subjects. One may elicit an anigraf by presenting subjects with a randomly chosen pair of alternatives (X, Y). A subject identifies a single feature that varies between the two alternatives (or indicates that no such feature exists).[7] The subject is first asked to extrapolate which alternatives, A and C, might be on either end of the pair and then to interpolate which alternative, B, is intermediate between the two items in the pair. Subjects may also indicate that there is no alternative

that is an appropriate interpolant or extrapolant (denoted by "|") or that there is an appropriate interpolant or extrapolant that is not listed in the set of alternatives (denoted by "O"). The result is an ordered quintuple of alternatives constituting a trajectory (i.e., a path) through an anigraf, (A,X,B,Y,C). This quintuple is then used to generate three triples: (A,X,B), (X,B,Y), and (B,Y,C). If the quintuple contains a | or a O as an extrapolant, it is truncated. For example (|,X,B,Y,C) and (O,X,B,Y,C) can only generate triples (X,B,Y) and (B,Y,C). Similarly, (A,X,|,Y,C) cannot generate any triples. On the other hand, transitivity suggests that (A,X,O,Y,C) can generate the triples (A,X,Y) and (X,Y,C).

Richards and Koenderink (1995) propose a set of rules for constructing anigrafs from triples—an approach that they have applied in the domains of human factors, mathematical psychology, and political science. Specific application domains include the analysis of musical intervals (Gilbert and Richards 1994); colors, textures, and geographic features (Richards and Koenderink 1995); travel routes (Lokuge et al. 1996); geometric shapes (Feldman and Richards 1998), and relationships between political movements (Richards 2001).

4.1 Statistical significance of anigrafs

Although Richards and Koenderink (1995) originally proposed using a triple's frequency as a continuous measure of how strongly structures are connected, this approach is highly sensitive to random variation. This paper therefore proposes a novel approach to constructing anigrafs from survey data. Assuming that preferences are unstructured

---

[7] A "feature" can, but need not, be a continuous design attribute. Indeed, taking continuous design attributes as given, and using them as features, can be misleading. For example, when trying to characterize the relationship between alternatives in a color space defined by hue, saturation, and lightness axes, Richards and Koenderink (1995) found that "transformations used by the subjects to move along a route in the constant lightness plane must vary in both saturation and hue simultaneously". Similarly, a feature may correspond to a transformation that takes alternative A into alternative B, as in Richards and Koenderink's (1995) texture space example, where they found that "the proper representation for textures is not simply a space of material types, but rather a space of types of transformations". In general, features capture the subject's perceptions of how various options are related or how one may be transformed into the other holding all else equal.

(i.e., unrestricted domain), it is straightforward to demonstrate that the probability that any randomly chosen triple appears m times is given by a Poisson distribution, $P(m) = \frac{\lambda^m e^{-\lambda}}{m!}$, with parameter $\lambda = T/N$ where T is the expected total number of triples elicited from experimental subjects and N is the number of unique triples that may be generated (in each case accounting for | and O entries and removing them as appropriate). A given triple is statistically significant if it appears significantly more often than would be expected per this Poisson distribution. Furthermore, an ensemble of triples constitutes a statistically significant anigraf at the $p < 0.05$ level if its family-wise error rate is less than 0.05. Thus, standard approaches to control family-wise error rate after multiple comparisons, such as the Holm-Bonferroni correction, apply.

4.2 Case study: smartphone anigraf

The ultimate purpose of this paper is to demonstrate how preference orders are constrained by empirical regularities such as similarity judgments. Thus, this section examines the relationship between anigraf data and preference judgments for the five Android smartphones listed in Table 4, adapted from Haston (2014). Similarity data for these five smartphones were collected from 36 experimental subjects who were recruited using a Human Intelligence Task (HIT) posted to Amazon's Mechanical Turk service on December 30, 2016. Each subject was asked to rank the similarity (on a scale of 0-100) of, and then to generate quintuples for, the ten pairs of smartphones in the list. 28 (78%) subjects completed all ten pairwise comparisons (data from the remaining eight subjects were excluded because they violated the same HIT instructions at least twice), yielding a total of 280 quintuples. These quintuples generated 840 triples, of which 517 (62%) were admissible (e.g., they did not possess a "|"). After completing the TM survey, subjects were also asked to rank the smartphones (indifference between options was allowed).

The associated survey protocol was determined to be exempt from IRB review by the George Washington University's Office of Human Research (IRB #061650; full survey protocol in the Supplemental Material). Six triples, each of which appeared at least 23 times, were retained, yielding a set that was statistically significant at the $p < 0.05$ level after adjusting for multiple comparisons using the Holm-Bonferroni correction (Table 5).

Although these triples significantly constrain the set of preference orders that may be considered (specifically, each anigraf is required to order alternatives as in the corresponding triples), Richards and Koenderink (1995) indicate that the rules for constructing anigrafs from TM data are not fully deterministic. Therefore, they suggest heuristics to select a best fitting anigraf given the set of triples (although see Gilbert 1997, for a simulated annealing algorithm that aims to address this limitation). Thus, the original trajectory mapping technique proposed by Richards and Koenderink (1995) is underspecified. Furthermore, "O" was entered by subjects for 296 (35%) of the 840 entries recorded, indicating that additional nodes may be missing from the set of five smartphones considered. In general, allowing subjects to volunteer "O"— indicating an intermediate category that is not included in the TM survey—allows for extra degrees of freedom in anigraf models. At one extreme, any time at least one subject indicates the presence of an "O", one could interpolate a new category to be included in the anigraf. Naturally, this would overfit the data to that subject's judgment. On the other extreme, one could ignore all "O" entries, relying on transitivity as mentioned above. This tends to underfit (or bias) the data to the limited set of categories suggested by the researcher. This is therefore an instance of the bias-variance tradeoff frequently encountered in model selection. Since the ultimate purpose of the analysis of anigrafs is to examine the extent to which similarity structure drives preferences, the next section presents an approach to model selection based on the

Table 4 Attributes of the top five Android smartphones reported by (Haston 2014)

	Nexus 5	LG G2	Samsung Galaxy S4	Samsung Galaxy J	HTC ONE
Display size	5.0" 1080p	5.2" 1080p	5.0" 1080p	5.0" 1080p	4.7" 1080p
Processor	Snapdragon 800	Snapdragon 800	Snapdragon 800	Snapdragon 800	Snapdragon 600
Camera	8MP	13MP	13MP	13MP	4MP
Storage/ RAM	16/32 GB, 2 GB	16/32 GB, 2 GB	16/32 GB, 2 GB	16 GB, 3 GB	16/32 GB, 2 GB
Battery	2300 mAh	3000 mAh	2600 mAh	2600 mAh	2300 mAh
Dimensions	137.9 × 69.2 × 8.6 mm	138.5 × 70.9 × 8.9 mm	136.6 × 69.8 × 7.9 mm	137 × 70 × 8.6 mm	137.4 × 68.2 × 9.3 mm
Weight	130 g	143 g	130 g	146 g	143 g

Table 5 Six triples elicited from anigraf survey data

Triple			Frequency	Possible interpretation
HTC One	Nexus 5	Samsung Galaxy S4	31	MP
HTC One	Nexus 5	Samsung Galaxy J	31	MP
LG G2	Samsung Galaxy S4	Samsung Galaxy J	25	(Battery OR display size)/RAM
HTC One	Nexus 5	LG G2	25	MP
Samsung Galaxy J	Samsung Galaxy S4	Nexus 5	23	Camera * RAM
LG G2	Samsung Galaxy J	Samsung Galaxy S4	23	Volume

Although each triple is associated with a possible interpretation (selected by the author), these triples were elicited from subjects independently of these interpretations. Other interpretations are possible, and this in no way affects the underlying anigraf structure

MP Megapixels

selection of a maximum-likelihood anigraf from a training set of preference data.

4.2.1 Anigraf model selection

Figure 4 summarizes the structures of 36 anigrafs, all of which are consistent with the triples in Table 5. This set is not collectively exhaustive; however, these selections are the most consistent with the triples collected. A training set consisting of an additional 81 preference orders was collected to adjudicate between these models. The test set constitutes the preferences expressed by the 28 subjects from whom the anigraf triples were derived. (Two of the preference orders in the training set and one of the preference orders in the test set were excluded because subjects expressed indifference for their top choice.) Pairwise comparisons for both the training and test sets were derived from these preference orders (Table 6). Logistic regression models were fit to the training data for each of the 12 anigrafs in Table 7 (downselected by inspection from the original set of 36) and a "null" model in which those who preferred a given phone were indifferent between all but their first choice options. Adjudication between these anigrafs proceeded as follows: given a pairwise comparison between two options, Y_1 and Y_2, the probability, $P(x)$, that a subject selected option Y_1 is given by $\ln\left(\frac{P(x)}{1-P(x)}\right) = ax$, where x is equal to $+1$ if a given anigraf predicted that Y_1 is preferred, -1 if Y_2 was preferred, and 0 if the anigraf predicted indifference. a, which indicates the strength of preference, was fit to each anigraf using standard L_2-norm regularization: i.e., a was selected to minimize the quantity $\sum_{x,Y_i}[\ln(1 + e^{-ax})] + a^2$. To avoid overfitting, standard goodness-of-fit metrics that penalize models with more parameters (Akaike Information Criterion; AIC, and Bayesian Information Criterion; BIC), were also calculated (Table 7). Here, the number of parameters in each model was given by the number of possible edges in the

associated anigraf: $\frac{n(n-1)}{2}$ edges, where n is the number of nodes. Finally, the log-likelihood of the test data was calculated for each anigraf. This same technique is easily extensible to the remaining 24 anigrafs (analysis omitted for brevity).

4.2.2 Results

Results were consistent across multiple methodologies. All anigrafs shown in Fig. 4 displayed similar gross preference structure, although as model likelihood and goodness-of-fit decreased, differences between actual and predicted preferences occurred more often. Even so, given the top choice expressed by each subject, 7 of the 13 (54%) models listed agreed with the data that the LG G2 was the top choice, and all models ranked the LG G2 among the top two choices. All models also ranked the HTC One as the worst choice and the Nexus 5 as second worst. None of the models expressed cyclic preferences. Finally, all of these models generated predictions for subjects' pairwise preferences that are within the test set's 95% confidence intervals (Table 6). Thus, consistent with Richards' modal hypothesis (Richards and Bobick 1988), these results indicate that preference orders and similarity rankings are strongly associated in this case. Consequently, it is reasonable to conclude that the axiom of unrestricted domain does not apply here.

Given subjects' top choices, predicted and actual numbers of subjects preferring each option in the Condorcet tally were almost perfectly correlated in the test set for the best fitting anigraf (number 12 in Table 7, corresponding to Fig. 4d), $r(8) = 0.99$, $p < 0.001$. Thus, the anigraf constrains the group-level preference order. Although preferences for the second- and third choice options (the Samsung Galaxy S4 and the Samsung Galaxy J) are reversed between this anigraf's predictions and the data, subjects' preferences between these two options were weak
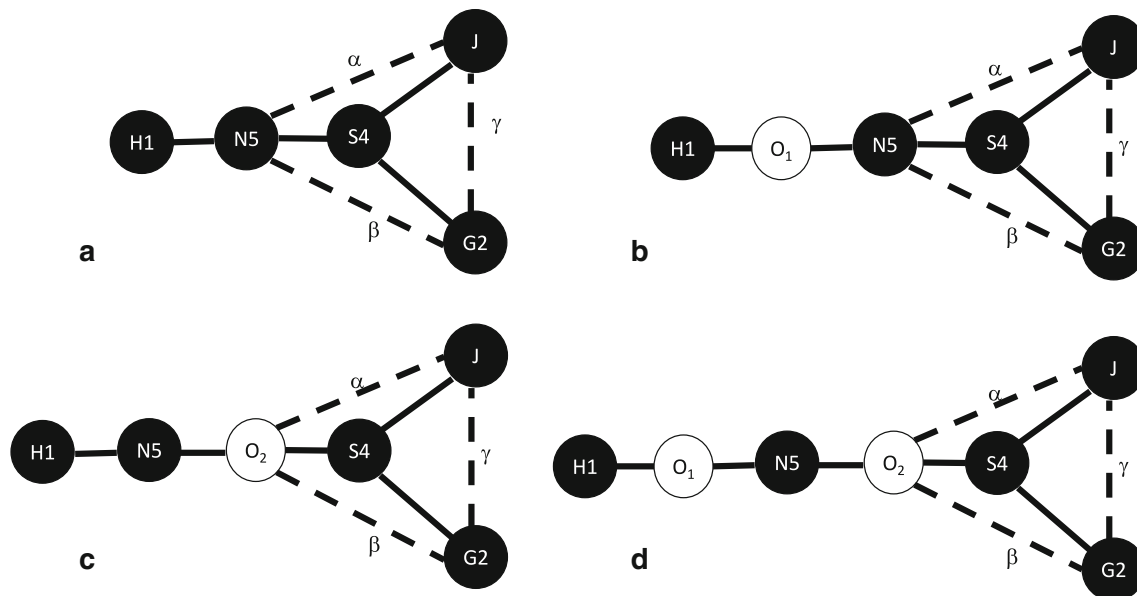
Fig. 4 Representation of 36 anigrafs that are consistent with the triples elicited from Trajectory Mapping survey data. *H1* HTC One, *N5* Nexus 5, *S4* Samsung Galaxy S4, *J* Samsung Galaxy J, *G2* LG G2. Since several subjects indicated "O" on their surveys, nodes O_1 or O_2 were included in several candidate anigraf models. Finally, edges α, β, and γ, although present in the elicited triples, need not be included because of the transitive property of similarity relations. The best fitting anigraf includes nodes O_1 and O_2, as well as edges α, β, and γ

(as shown by the small magnitudes in the both the test set and predicted Condorcet tallies in Table 6), suggesting that this preference reversal is a consequence of measurement error. Furthermore, independent of the test data, this same anigraf had the lowest AIC and BIC values, indicating that this model's fit to the training data is parsimonious—i.e., the extra model complexity resulting from adding two extra nodes that were not on the initial list of five smartphones is nevertheless worth the additional explanatory power. Finally, pairwise distance between nodes is strongly associated with average pairwise similarity rankings (shown in Table 6) elicited from the 28 subjects in the test set, $r(8) = 0.90$, $p < 0.001$.

Beyond these findings, several anigrafs that incorrectly predicted indifference between two options (such as anigraf 3 in Table 7, which predicts that those who prefer the Samsung Galaxy S4 will be indifferent between the Nexus 5 and the Samsung Galaxy J), or that predicted a strong preference when data indicated that there was none (such as anigrafs 1 and 8, which predict that those who prefer the Samsung Galaxy J will prefer the Samsung Galaxy S4 over the LG G2) were ruled out by model selection. Remaining deviations from this best fitting model (i.e., differences in preference whose magnitudes were not sufficiently large as to be statistically significantly) were roughly symmetrically distributed as predicted. Naturally, recruitment of additional subjects would allow one to capture even more nuanced anigraf structure.

4.2.3 Empirical assessment of knowledge depth

The model in Sect. 3 demonstrates that even minimal constraints on preference orders are sufficient to greatly reduce the likelihood of top cycles. The smartphone TM data allow for an empirical assessment of the knowledge depth parameter in that case. Specifically, the best fitting anigraf can be used to assess the relationship between strength of preference and knowledge depth. Strength of preference was measured as the absolute value of the normalized entries in the Condorcet tally (for example, if 9 out of 10 subjects preferred option Y_1 and the remaining subject preferred option Y_2, strength of preference was 80%; if subjects were split such that 5 subjects preferred option Y_1 and 5 preferred option Y_2, strength of preference was 0, etc.) and knowledge depth was measured as the minimum distance between the decision-maker's topmost preferred option and the options being compared. These two quantities were significantly anticorrelated, $r(48) = -0.51$, $p < 0.001$. A linear fit of these data is given by the equation $d = -1.4s + 1.99$ where s is strength of preference and d is the average distance between the node performing the evaluation and the closest node being evaluated. This equation indicates that when decision-makers assess options represented by nodes that are two hops away or more, subjects are, on average, indifferent between these options suggesting a knowledge depth of unity for this task.

Table 6 Subjects' pairwise similarity judgments and preferences

1st choice	N5				G2			S4		J
2nd choice	G2	S4	J	H1	S4	J	H1	J	H1	H1
Subjects' pairwise similarity judgments										
Mean (SD)	50.14 (21.44)	32.18 (18.82)	48.96 (21.02)	51.68 (23.05)	28.82 (19.85)	33.21 (21.98)	67.43 (20.88)	20.82 (21.01)	61.86 (22.20)	60.64 (23.07)
1st choice (*n*)	*Training set Condorcet tally*									
N5 (9)	9	9	9	9	3	3	3	3	−1	−3
G2 (22)	−22	−18	−14	21	22	22	22	13	20	20
S4 (12)	−6	−12	−8	10	−12	−2	10	12	12	10
J (6)	−6	−5	−6	6	−2	−6	6	−6	6	6
H1 (3)	−1	1	1	−3	1	3	−3	0	−3	−3
1st choice (*n*)	*Test set Condorcet tally*									
N5 (1)	1	1	1	1	−1	−1	−1	−1	1	1
G2 (12)	−12	−10	−10	12	12	12	12	9	12	12
S4 (4)	−4	−4	−4	4	−4	0	4	4	4	4
J (9)	−5	−7	−9	9	−1	−9	9	−9	7	9
H1 (1)	1	1	1	−1	−1	−1	−1	−1	−1	−1
Total	−19	−19	−21	25	5	1	23	2	23	25
95% CI	[−9.46, −23.81]	[−9.46, −23.81]	[−11.85, −24.92]	[26.65, 17.13]	[13.76, −5.01]	[10.40, −8.65]	[25.89, 14.38]	[11.47, −7.85]	[25.89, 14.38]	[26.65, 17.13]
1st choice (*n*)	*Condorcet tally predicted by best fitting anigraf given test set's top choices*									
N5 (1)	1	1	1	1	0	0	0	0	0	0
G2 (12)	−12	−12	−12	12	12	12	12	0	12	12
S4 (4)	−4	−4	−4	4	−4	0	0	0	0	0
J (9)	−9	−9	−9	9	0	−9	9	−9	9	9
H1 (1)	1	1	1	−1	0	−1	0	−1	−1	−1
Total	−23	−23	−23	25	8	3	24	−5	24	24

Each entry in the Condorcet Tally sections indicates the difference between the number of subjects preferring the first option and the number of subjects preferring the second option, with positive numbers indicating an overall preference for the first option. For example, the entry corresponding to a pairwise comparison between S4 and J by the 22 subjects listing the J as their top choice has a value of 13 in the Training Set because 16 subjects preferred the S4, 3 subject preferred the J, and 3 subjects were indifferent (16 − 3 = 13)

N5 Nexus 5, *G2* LG G2 *S4* Samsung Galaxy S4, *J* Samsung Galaxy J, *H1* HTC one, *SD* standard deviation, *n* number of subjects listing the smartphone in the leftmost column as their first choice, *CI* confidence intervals

5 Discussion

The analysis presented here shows that the theoretically motivated possibility of cyclic preferences is an important constraint on decision-based design that cannot be ruled out entirely. However, even minimal domain restrictions can vastly reduce the likelihood of cyclic preferences. Thus, Arrow's Theorem, which has previously been characterized as widely requiring "dictatorship" in engineering design teams with multiple stakeholders, does not usually apply.

Without domain restrictions, the probability of cyclic preferences swiftly goes to unity when there are more than three alternatives. However, this assumes that all cyclic preferences, not just top cycles, are problematic. Our results show that the probability of top cycles remains less than 50% if the number of alternatives remains small (12 or fewer). However, when the number of alternatives becomes

Table 7 Goodness of fit metrics for 10 anigraf models

	Nodes	Edges	a	AIC	BIC	LL	r	Preference order
Null			3.37	444.80	444.31	−121.18	0.78	G2 > S4 > J > N5 > H1
1	None	None	2.14	392.65	412.16	−73.64	0.95	S4 > G2 > J > N5 > H1
2	None	γ	2.34	377.14	396.65	−71.58	0.99	G2 > J > S4 > N5 > H1
3	None	α, β, γ	1.63	558.89	578.41	−134.73	0.67	G2 > J > S4 > N5 > H1
4	O_1	β	2.23	364.70	393.93	−65.65	0.98	S4 > G2 > J > N5 > H1
5	O_1	γ	2.13	393.83	423.10	−66.61	0.99	G2 > S4 > J > N5 > H1
6	O_1	α, β	2.33	373.75	403.02	−61.02	0.97	S4 > G2 > J > N5 > H1
7	O_1	α, β, γ	2.37	377.05	418.03	−61.11	0.99	G2 > J > S4 > N5 > H1
8	O_1, O_2	None	2.15	391.32	432.30	−63.82	0.98	S4 > G2 > J > N5 > H1
9	O_1, O_2	β	2.44	348.52	389.49	−62.74	0.99	S4 > G2 > J > N5 > H1
10	O_1, O_2	γ	2.26	390.43	431.40	−62.00	0.99	G2 > J > S4 > N5 > H1
11	O_1, O_2	α, β	2.68	340.22	390.20	−60.59	0.97	S4 > G2 > J > N5 > H1
12	**O_1, O_2**	**α, β, γ**	**2.73**	**339.50**	**380.47**	**−54.16**	**0.99**	**G2 > J > S4 > N5 > H1**

The best fitting model, number 12, is shown in bold. A "null" model, in which subjects prefer only their top choice, and are otherwise indifferent, is included for reference

a estimated parameter of logistic regression model associated with this anigraf, *AIC* Akaike Information Criterion of training set, *BIC* Bayesian Information Criterion of training set, *LL* log-likelihood of test set, r Pearson correlation between predicted and actual Condorcet tallies on test set, *H1* HTC One, *N5* Nexus 5, *S4* Samsung Galaxy S4, *J* Samsung Galaxy J, *G2* LG G2

large even the probability of top cycles goes to unity. Since modern engineered systems must typically explore vast trade spaces with very large numbers of options, restricting one's attention only to unstructured preferences with an unrestricted domain is no guarantee of a winning outcome.

The assumption of unrestricted domain is not empirically justifiable given the vast literature on mental models in engineering design. Following Richards et al. (1998; 2002), the analysis presented in this paper shows that cyclic outcomes can be avoided in the vast majority of cases if decision-makers can agree on similarity relationships between design alternatives, using a simple technique relying on Condorcet tallies and pairwise comparisons. This need not in any way restrict their choice of their most preferred alternative. Furthermore, total agreement on the knowledge structure is not necessary. When decision-makers' mental models minimally overlap ($k = 1$), the probability of a top cycle falls below 5%, even for 100 alternatives. Furthermore, a knowledge depth of $k = 3$ yields results that are virtually indistinguishable from fully constrained preference orders when the number of alternatives exceeds 20.

Data regarding similarity relations between alternatives can be extracted in a straightforward manner using techniques such as TM. Since individual designers indicate their preferences independently, groupthink need not dominate. However, once an anigraf has been generated from the data gathered from all stakeholders, these elicited results can encourage convergence on a common mental model, further increasing the likelihood of a rational aggregate preference order. Since the results of the

Condorcet tally are guaranteed to be optimal (Young 1995), this approach need not sacrifice internal consistency to obtain empirical correspondence. Finally, once an anigraf has been elicited, it can be analyzed to determine if the structure could lead to top cycles, prior to the commitment of significant resources. Only in these rare situations may a single decision-maker be necessary. Thus, the novelty of this approach is that it enables the systematic incorporation of a much wider range of perspectives in design, maintaining both consistency with an axiomatic framework while making the correspondence with empirical regularities explicit. When these regularities are physical in nature, designers will have better information regarding the extent to which their preferences might be constrained. Moreover, when these regularities are socio-cultural, this information may help designers to better understand the extent to which their values limit possible design options.

5.1 Comparisons to other models

The results of the simulation presented here are not restricted to social choice problems. If one chooses to interpret the preferences over nodes in an anigraf as preferences over options imposed by (potentially single-peaked) design criteria, consistent with Franssen's (2005) argument, the same logic applies to engineering design requirements. Here, relationships between the preferences orders imposed by each design criterion in isolation are structured by physical laws and other regularities (e.g., Broniatowski and Weigel 2008, discuss regular relationships between the political and technical domain in the

context of human space exploration). Specifically, the ordering over design options imposed by these criteria constrain one another to the extent that they are correlated or related in a lawful manner, such that a high-dimensional tradespace may be approximated by a lower-dimensional subspace. Eliciting these correlations can be straightforward when design criteria are physically well-characterized. Furthermore, even if one does not know a given design's criteria *a priori* (precluding one's ability to guarantee with absolute certainty that cyclic outcomes are avoidable), simulation results indicate that top cycles are the exception rather than the rule. On the other hand, designers have not yet come to consensus regarding how to measure non-traditional requirements, such as resilience, adaptability, etc. Examining the relationship between cyclic outcomes and non-traditional design requirements is therefore a topic that would benefit significantly from future empirical research.

Franssen and Bucciarelli (2005) state that broadening the concepts of rationality and optimality in engineering design is a valuable goal, noting that existing concepts from the social choice literature have direct bearing on design decision problems and thus deserve more attention. Specifically, Franssen and Bucciarelli (2005) argued that the definition of rationality in design decision-making should be expanded to account for multiple stakeholders who do not share a common utility function. Contrary to Hazelrigg (1997), they note that "…the possibilities for [a] single individual calling the shots, [or] setting-out a single utility function which would yield an optimum, appear slim." (p. 949), motivating the use of game theory to select a design. Illustrative examples such as the classic Prisoner's Dilemma, show that, absent cooperation, such approaches are not Pareto-optimal (meaning that they do not maximize the total utility when summed across all stakeholders). Thus, Franssen and Bucciarelli advocate for negotiation among stakeholders, stating that "If they are rational and if they have the possibility of arranging a form of cooperation that enables them to realize the design jointly, rationality will prompt them to do so" (p. 949)—an approach that is broadly consistent with the one presented in this paper. However, there are important differences. Franssen and Bucciarelli's approach aims to find the Pareto-optimal solution from among a set of individual utility functions such that all stakeholders agree on outcomes. This may not be achievable in the event of utility functions that are at odds with one another, as in zero-sum games. In contrast, the approach presented here demonstrates that it may indeed be possible to define a group-level preference ordering. This approach does not require agreement on outcomes, or even on preferences; rather, it requires only a minimal degree of overlap between mental models. Furthermore, this overlap is demonstrably optimal given

empirical limitations (per Condorcet's formulation, discussed by Young 1995). Thus, the technique presented here refutes Hazelrigg's (1996) assertion that "any methodology that demands the construction of a group utility function in any aspect of its construct is logically inconsistent and doomed to failure" (p. 162).

Others might object that heuristic approaches, such as PuCC, have strengths that this approach does not. For example, Frey et al. (2009) emphasized the role of ideation using PuCC, whereas the approach presented here does not explicitly allow for the incorporation of new alternatives in the middle of the decision-making process. However, the construction of anigrafs can be construed as an iterative process. During each iteration, nodes can included or excluded based upon prior survey results, the collection of pilot data, etc. Between these iterations, ideation techniques may be used to generate alternatives for inclusion in the anigraf surveys. The strength of the approach presented here is that it combines axiomatic coherence with empirical correspondence in a principled way; nevertheless, the role of ideation in design remains an important topic for future work.

On the other hand, the axiomatic limitations of the PuCC are well known (Hazelrigg 2010). This paper does not claim that all, or even most, heuristic techniques can avoid cyclic preferences. The literature, and especially the work of Hazelrigg, has taken for granted that cyclic preferences are unavoidable necessitating "dictators". If, however, there exists even a single method that can reliably avoid cyclic preferences in most cases, and if it is possible to determine, in advance, when cyclic preferences can be ruled out empirically, then that method could be used in concert with existing heuristic techniques to avoid cyclicality.

5.2 Limitations and directions for future work

The techniques discussed in this paper are premised on the idea that similarity of decision alternatives implies similarity of preference. In other words, one's second choice will be the option that is most similar to one's top choice. This assumption is widely used in multiple domains; for example, Richards (2001) gathered data showing that preferences within domains as separate as political choice and movie rentals conformed to this model. Our empirical results are consistent with this assumption, and there is no reason to believe that it would be violated within the domain of engineering design; nevertheless, more extensive empirical validation of this framework remains an immediate next step for future work.

Similarly, one might object that preferences over alternatives do not necessarily correspond to preferences over outcomes. This is the essence of Franssen's (2005) critique

of Scott and Antonsson (1999). This paper addresses this objection using Richards and Bobick's (1988) observation that mental constructs are themselves shaped by empirical regularities. Furthermore, the feasible regions of trade-spaces are highly structured by regularities that are at least as strong. To the extent that outcomes depend on alternatives, preferences should follow suit. Nevertheless, one might further object that outcomes themselves cannot be anticipated in sufficiently complex situations. This is a general critique of decision-based design that is not limited to the techniques discussed in this paper—indeed, this critique has been leveled against utility theory in general. Addressing this critique is outside of the scope of this paper, since it directly contradicts the foundations of any axiomatic approach.

Finally, one might object that the framework presented in this paper does not account for design teams whose members have stark differences in their mental models. Specifically, these models may sharply diverge if participants hold very different ways of structuring world information. In such circumstances, cyclic preferences may indeed dominate. For example, Bucciarelli (1994), notes that professional disciplines function as separate "cultures", which could lead one to conclude that multidisciplinary design teams are especially likely to be subject to cyclic preferences. However, recent data indicate that members of multidisciplinary design teams do indeed construct shared mental models (Avnet 2015, 2016; Avnet and Weigel 2013) suggesting that culture, as defined by Bucciarelli, does not preclude shared cognitive representations (see also, Romney et al. 1986, 1996). Future work should therefore focus on determining the situations under which group members' mental models diverge in a manner that allows for cyclic preferences to occur.

In such circumstances, techniques such as problem structuring methods (PSMs; e.g., Mingers and Rosenhead 2004) may be instrumental in establishing a common mental representation prior to selection of an alternative for the group. Furthermore, the techniques presented here allow for the evaluation of the success of these methods by measuring the extent to which mental models overlap before and after these methods are used. Indeed, anigraf data could be elicited in concert with these methods, first by helping decision-makers to come to some agreement regarding what alternatives should be included, and then by helping decision-makers to construct a common mental model. This, in turn, requires extension and refinement of the empirical techniques discussed in this paper.

One may further object that constructing an anigraf is labor intensive due to the large number of pairwise comparisons needed to fully construct a preference order. Indeed, prior authors (Frey et al. 2009) have commented on the associated workload (although Dym, Wood, and Scott

2002, note that pairwise comparisons are "cheap and require little detailed knowledge, and are thus valuable in conceptual design", p. 241). Furthermore, the statistical approach presented here, based on the Poisson distribution, can be used to perform a power analysis to determine the total number of pairwise comparisons necessary to establish a meaningful structure. Thus, it is conceivable that not all subjects need to perform all pairwise comparisons, especially since only minimal structural constraints are necessary to severely limit the likelihood of cyclic preferences. Furthermore, this analysis suggests a compelling new research agenda for decision-based design: the empirical elicitation of mental models. Although some work has been performed in this area (e.g., Ahamed et al. 2016; Doyle and Ford 1998; Moray 1990; Sterman 1994) little attention has been devoted to the specific interaction between mental model formation and design decision-making. A full treatment of this topic is left to future work.

Finally, this paper does not address the question of when group members should give up trying to find a common design, such as when their preferences are so divergent that the best outcome at the level of the group is nevertheless of very low value to some, or all, group members. Although outside the scope of this paper, this question is related to the issue of how to design flexibility into a system such that it can change its functionality as needed to support multiple decision-makers, environments, etc. (Broniatowski 2017). This flexibility often leads to quite complex designs and/or several rework cycles, although some system architectures handle this tradeoff better than others (Broniatowski and Moses 2016). An intriguing direction for future work would combine this stream of research with the techniques presented in this paper.

6 Conclusion

In conclusion, there are several novel contributions of this paper. Its primary contribution is to demonstrate that adherence to design axioms need not conflict with empirical correspondence. The analysis presented here reduces the debate between proponents of scientism and praxis to an empirical question: *For this design*, does the structure of the commonly held knowledge allow cyclic preferences? Similarly, for multi-criteria decision problems, this question may be formulated as follows: *For this design* do the relationships between design criteria allow cyclic preferences?

In the theoretical domain, this paper is the first to examine the extent to which the axiom of unrestricted domain applies to engineering design outside of the restrictive context of single-peaked preferences—the only domain restriction considered in the engineering design

literature in any depth. Specifically, using a novel simulation technique that builds upon the pioneering work of Richards and colleagues (Richards et al. 2002; Richards 2015) this paper demonstrates that single-peaked preferences are overly restrictive. Several different anigrafs possess structures that are sufficient to avoid top cycles.

In the empirical domain, this paper proposes the use of a technique that has previously not been applied to engineering design—Trajectory Mapping—to measure the extent to which empirical data corresponds to axiomatic coherence. This technique is certainly not the only one that can accomplish these goals; however, it does indicate that such a combination is both plausible and feasible. In addition, this paper is the first to relate the work on mental models in engineering design to preference orders, thus connecting two previously disparate bodies of literature.

This work also represents several computational advances. Specifically, it is the first to use a statistical technique, based upon the Poisson distribution, to elicit similarity triples from anigraf extrapolant and interpolant data. It is also the first to use a model selection approach, based upon binomial logistic regression, to adjudicate between multiple anigrafs that are consistent with elicited triples. Specifically, a training dataset is used to select a best fitting anigraf whose predicted preference orders are then compared against a test set of preferences. Using this technique, which is applied here for the first time to a multi-criterion decision problem by a group in a technical domain (selection among smartphones), this paper has empirically verified that similarity relations are associated with preference orders, and that strength of preference varies with knowledge depth.

In conclusion, this paper demonstrates that coherence and correspondence need not be opposing imperatives in design. Indeed, an ideal approach would combine elements of both. Empirical information may be used to augment the correspondence of decision-based design methods while still maintaining axiomatic coherence.

References

Ahamed AB, Mohanty SD, Avnet MS (2016) A systems perspective on mental models research: toward a shared understanding of team cognition. In: 2016 conference on systems engineering research, Huntsville, AL

Anderson JR (1995) Cognitive psychology and its implications. Freeman, New York

Arrow KJ (1963) Social choice and individual values. Yale University Press, New Haven

Avnet MS (2015) A network-based approach to organizational culture and learning in system safety. Procedia Comput Sci 44:588–598

Avnet MS (2016) A network-based analysis of team coordination and shared cognition in systems engineering. Syst Eng 19:395–408

Avnet MS, Weigel AL (2013) The structural approach to shared knowledge: an application to engineering design teams. Hum Factors 55:581–594

Bang M, Medin DL, Atran S (2007) Cultural mosaics and mental models of nature. Proc Natl Acad Sci 104:13868–13874

Barberà S, Gul F, Stacchetti E (1993) Generalized median voter schemes and committees. Journal of Economic Theory 61:262–289

Black D (1948) On the rationale of group decision-making. J Polit Econ 56:23–34

Broniatowski DA (2017) Flexibility due to abstraction and decomposition. Syst Eng. doi:10.1002/sys.21381

Broniatowski DA, Moses J (2016) Measuring flexibility, descriptive complexity, and rework potential in generic system architectures. Syst Eng 19:207–221. doi:10.1002/sys.21351

Broniatowski DA, Weigel AL (2008) Articulating the space exploration policy–technology feedback cycle. Acta Astronaut 63:649–656

Bucciarelli LL (1994) Designing engineers. MIT press, Cambridge

de Caritat marquis de Condorcet MJAN (1785) Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. L'imprimerie royale

Doyle JK, Ford DN (1998) Mental models concepts for system dynamics research. Syst Dyn Rev 14(1):3–29

Dym CL, Wood WH, Scott MJ (2002) Rank ordering engineering designs: pairwise comparison charts and Borda counts. Res Eng Des 13:236–242. doi:10.1007/s00163-002-0019-8

Feldman J, Richards W (1998) Mapping the mental space of rectangles. Percept Lond 27:1191–1202

Franssen M (2005) Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. Res Eng Des 16:42–56

Franssen M, Bucciarelli LL (2005) On rationality in engineering design. J Mech Des 126:945–949. doi:10.1115/1.1803850

Frey DD, Herder PM, Wijnia Y, Subrahmanian E, Katsikopoulos K, Clausing DP (2009) The Pugh controlled convergence method: model-based evaluation and implications for design theory. Res Eng Des 20:41–58

Gilbert SAB (1997) Mapping mental spaces: How we organize perceptual and cognitive information. Massachusetts Institute of Technology Unpublished Dissertation

Gilbert SA, Richards W (1994) Using trajectory mapping to analyze musical intervals. In: Proceedings of the sixteenth annual conference of the Cognitive Science Society

Haston S (2014) Top five 5-inch smartphones compared. In: Mobile Geeks. http://www.mobilegeeks.com/top-5-inch-smartphones-compared/. Accessed 8 July 2016

Hazelrigg GA (1996) The implications of arrow's impossibility theorem on approaches to optimal engineering design. J Mech Des 118:161–164. doi:10.1115/1.2826864

Hazelrigg GA (1997) On irrationality in engineering design. J Mech Des 119:194–196. doi:10.1115/1.2826235

Hazelrigg GA (1998) A framework for decision-based engineering design. J Mech Des 120:653–658. doi:10.1115/1.2829328

Hazelrigg GA (1999) An axiomatic framework for engineering design. J Mech Des 121:342–347. doi:10.1115/1.2829466

Hazelrigg GA (2010) The Pugh controlled convergence method: model-based evaluation and implications for design theory. Res Eng Des 21:143–144. doi:10.1007/s00163-010-0087-0

Katsikopoulos KV (2009) Coherence and correspondence in engineering design: informing the conversation and connecting with judgment and decision-making research. Judgm Decis Mak 4:147

Katsikopoulos KV (2012) Decision methods for design: insights from psychology. J Mech Des 134:084504

Langan-Fox J, Code S, Langfield-Smith K (2000) Team mental models: techniques, methods, and analytic approaches. Hum Factors 42(2):242–271

Langan-Fox J, Anglim J, Wilson JR (2004) Mental models, team mental models, and performance: process, development, and future directions. Hum Factor Ergon Man 14(4):331–352

Lokuge I, Gilbert SA, Richards W (1996) Structuring information with mental models: a tour of boston. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, USA, pp 413–419

Mathy F, Feldman J (2012) What's magic about magic numbers? Chunking and data compression in short-term memory. Cognition 122:346–362

Mingers J, Rosenhead J (2004) Problem structuring methods in action. Eur J Oper Res 152:530–554

Miller G (1956) Human memory and the storage of information. IRE Trans Inform Theor 2:129–137

Moray N (1990) A lattice theory approach to the structure of mental models. Philos Trans R Soc Lond B 327:577–583

Pugh S (1991) Total design: integrated methods for successful product engineering. Addison-Wesley, Reading

Reich Y (2010) My method is better! Res Eng Des 21:137–142

Reyna VF, Lloyd FJ (2006) Physician decision making and cardiac risk: effects of knowledge, risk perception, risk tolerance, and fuzzy processing. J Exp Psychol Appl 12:179

Richards D (2001) Coordination and Shared Mental Models. Am J Polit Sci 45:259–276. doi:10.2307/2669340

Richards W (2008) Modal Inference. In: Proceedings of the AAAI symposium on "naturally inspired artificial intelligence"

Richards W (2015) Anigrafs: experiments in cooperative cognitive architecture. The MIT Press, Cambridge

Richards W, Bobick A (1988) Playing twenty questions with nature. In: Pylyshyn ZW (ed) Computational processes in human vision: an interdisciplinary perspective. Ablex Publishing Corp., Norwood, pp 3–26

Richards W, Koenderink JJ (1995) Trajectory mapping: a new nonmetric scaling technique. Perception 24:1315–1331

Richards D, McKay BD, Richards WA (1998) Collective choice and mutual knowledge structures. Advs Complex Syst 01:221–236. doi:10.1142/S0219525998000156

Richards W, McKay BD, Richards D (2002) The probability of collective choice with shared knowledge structures. J Math Psychol 46:338–351

Romney AK, Weller SC, Batchelder WH (1986) Culture as consensus: a theory of culture and informant accuracy. Am Anthropol 88:313–338

Romney AK, Boyd JP, Moore CC, Batchelder WH, Brazill TJ (1996) Culture as shared cognitive representations. PNAS 93:4699–4705

Saari DG (1994) Geometry of voting. Studies in economic theory, vol 3. Springer, New York

Saari DG, Sieberg KK (2004) Are partwise comparisons reliable? Res Eng Des 15:62–71

Saaty TL (1990) How to make a decision: the analytic hierarchy process. Eur J Oper Res 48:9–26

Scott MJ, Antonsson EK (1999) Arrow's theorem and engineering design decision making. Res Eng Des 11:218–228

Sterman JD (1994) Learning in and about complex systems. Syst Dyn Rev 10(2):291–330

Young P (1995) Optimal voting rules. J Econ Perspect 9:51–64