FROM CONSTRAINED DESIGN TO DESIGNED CONSTRAINTS TO CONSTRAINT REVERSAL

Thomas Fischer¹ and Laurence D. Richards²
¹Department of Architecture, Xi’an Jiaotong-Liverpool University, Suzhou, China, Thomas.Fischer@xjtlu.edu.cn
²Indiana University East, Richmond, Indiana, USA laudrich@iue.edu

ABSTRACT:
This paper traces the changing notion of constraints in design along with the changing notion
of systems since the mid-20th century in the intersection of design theory and systems
theory. Taking a second-order cybernetic perspective, the paper develops constraints as
observer dependent, and it analyses conditions under which constraints tend to be beneficial
or detrimental. Ethical implications of constraints in design processes are established with
reference to system boundaries. Constraint-oriented design is discussed as an alternative to
goal-oriented design, and finally a method named constraint reversal is introduced as a
strategy of deliberate defiance of constraints to support design exploration.

1. INTRODUCTION: CONSTRAINT AS A SYSTEMS CONCEPT

Amongst the first to explicitly address the role of constraints in design was trained
mathematician and systems analyst Horst Rittel, professor and head of the Ulm School of
Design (HfG) and later professor at the University of California, Berkeley. Probably best
known for the distinction between “wicked” and “tame” problem that he proposed together
with Webber (Rittel and Webber 1973), Rittel was also amongst the first to apply a “systems”
approach to design. Rittel is thus an early representative of thinkers in the intersection
design theory and systems theory. Second-order cybernetic design theory is a
contemporary development of this intersection, represented for example by Glanville (1999),
Jonas (2007) and Krippendorff (2006). In the following, cybernetics and the development of
the second-order cybernetic perspective will be taken as a vantage point from which to
examine the role of constraints in design.

2. CONSTRAINT AND VARIETY

Rittel frequently referred to cybernetician W. Ross Ashby (1956) and his work on variety.
Based on the notion of variety, Rittel proposed a fundamental division of design activities
that remained central to his teaching throughout his career (Protzen and Harris 2010 p.
Ashby’s work on variety in cybernetics gave prominence to the notion of constraint (Ashby
1956 p. 127ff.), which he defined as a “relation between two sets, [which] occurs when the
variety that exists under one condition is less than the variety that exists under another”
(ibid.). As an example Ashby offers (British) traffic lights, which are constrained so that only
four combinations of the on/off states of red, yellow, and green occur (one condition),
whereas eight such combinations would be available (other condition) if the three on/off
states could vary independently. Rittel then defines constraints as “certain values of the
decision variables or design variables [that] are excluded”. Using a spatial metaphor for
variety he continues: “[a]ll the knowledge the designer has, of a scientific or professional
character, can be plotted into the decision space as constraints, for example a ‘maximum
weight 5 tons’ constraint would cut off all solutions above that figure”.

Important distinctions have to be drawn between the traffic lights example and the decision
not to consider weights above some threshold in a design process. One is the distinction
between mechanisms and human decision making. Another is the distinction between
processes that are constrained from the outside and processes that are constrained from the
inside (Glanville 1997). This echoes Miller (1978 p. 659) who notes that “[c]ertain alternative
solutions are eliminated for an organization’s decider by constraints on it, either from inside
the system or from outside”. As will be shown below, the question of what resides within or
outside of a system depends on the understanding of the term system, and, according to the second-order cybernetic understanding, on the system’s observer. The traffic lights are a mechanism which is constrained from the outside, by traffic rules and those who implement traffic lights. The design decision to consider weights below 5 tons only is made by a human acting within the decision process in question. The mechanism/human distinction and the inside/outside distinction are related in as far as humans appear to be more capable than mechanisms to engage appositely with what lies outside their systemic boundaries.

From positivist origins, both cybernetic theories and Rittel’s theories of design have evolved over time. To develop a differentiated view of constraints in design, it is necessary to quickly trace the changing understanding of the systems notion (including ‘systems boundary’, which is essential to the inside/outside distinction drawn above) during the second half of the past century.

3. SYSTEMS, OBSERVERS AND BOUNDARIES

As an early definition, Bertalanffy (1968 p. 38) describes the term system as a “(set) of elements standing in interrelation”. This notion was taken up by early design researchers (Gosling 1963 p. 24) who initially tended to apply the term system to outcomes of design, i.e. to parts or components making up products. In this view, which is still encountered in systems engineering and the engineering design fields, constraints are frequently used to refer to limits imposed on variable elements of products, such as moving parts.

Others further differentiate Bertalanffy’s definition by recognizing the existence of systems in contexts or environments, and the interaction of systems with their environments via inputs and outputs that cross a given system’s boundaries. Further differentiations ascribe more characteristics to systems, including teleological, i.e. goal-oriented (Churchman 1971 p. 42) and self-perpetuating (Maturana and Varela 1987 p. 47) ones. Weinberg (2001 pp. 51ff.) introduces a subjective, observer-dependent aspect by defining system as “[a] way of looking at the world”. Rittel (1992 p. 59), analogously, states that a system “reflects someone’s understanding of something” (“Ein System reflektiert jemandes Verständnis von etwas”).

Rittel’s recognition of systems as observer-dependent is one aspect of his proposal of a second-generation of the systems approach, of design theory and of design methods, based on his and others’ initial development, and later rejection of more positivist and more determinist “first generation” ones (Cross 1984 pp. 317-327). This development is somewhat analogous to the recognition of observer dependence (amongst other aspects) in cybernetics, i.e. the development of second-order cybernetics. As an aside, it should be mentioned that there are other notable analogies between Rittel’s move from first-generation to second-generation design theories and the move from first order to second-order cybernetics. Rittel’s description of the designerly condition as a “symmetry of ignorance” (Cross 1984 p. 325; Protzen and Harris 2010 p. 159), for example, resembles the symmetry of conversations (Pask 1975) and the value of ignorance (besides knowing) noted in second-order cybernetic design theory (Glanville 2004 pp. 423ff.).

In second-order cybernetics, systems are understood as bounded by the way observers distinguish them. The systems notion, and along with it the systems boundary notion, thus becomes subjective, corresponding to Spencer-Brown’s dictum “Draw a distinction” (Spencer-Brown 1997). Brün (2004 p. 82), accordingly, states: “A system is not something that exists objectively in space or time anywhere. A system is the result of a look at a collection of stipulated elements. Stipulated in that I say which elements I will look at.” Instead of phrases like “defining a system” or “specifying a system” some second-order cyberneticians hence use the phrase “looking a system”.

One may, for instance, choose to consider a person, say a child, as a system in the sense that the child is an organism made up of organs (a set of elements standing in interrelation). Alternatively, one may choose to recognize the child’s dependency on her mother, without whose care and feeding the child would not survive. Now, the child and (at least aspects of) the mother form a system. The observer has the freedom and responsibility to choose what is relevant and considered as parts of a “looked” system. Analogously, a designer may
choose to approach a design challenge at the level of an apartment building, or, realizing some broader implications and dependencies, at the scale of an apartment block or a city block. Rittel (1973) explains: "[e]very wicked problem can be considered to be a symptom of another problem" (ibid. p. 165), and "[t]he choice of [a wicked problem's] explanation determines the nature of the problem's resolution" (ibid. p. 166).

In this view, "system" is understood as whatever set of elements an observer considers as acting together—for example, following a common goal (a teleological approach). Oftentimes, systems we look (at) are bounded by physical skins, coatings, housings or some similar demarcation of a boundary. The above-mentioned child, for example, has skin and clothing, cells have membranes, birds’ eggs have shells, computers have cases, buildings have façades, communities use buildings and so on. This can be explained by taking these systems as having “evolved” or developed, or as having been given armor and structure to protect them from destructive forces so as to maintain their coherence and integrity. The skin of a child may likely coincide with the boundary an observer projects onto the child system seen as an autonomous entity, because the forces that brought forth human skin, themselves acting as observers, may have also approached the human body as an autonomous entity. If, however, an observer chooses not to see the child as autonomous but as dependent on her mother, then that observer is free to project a boundary which intersects with the boundaries of the mother. This boundary now does not coincide with any physical hull. It is a projection brought forth by an observer’s way of looking. This projected boundary, regardless of whether it coincides with a physical boundary or not, is the systems boundary as approached by second-order cybernetics. In this way, in the second-order cybernetic view, systems and their boundaries are observer-dependent, and in principle independent from physical configuration (Fischer 2012 pp. 679-680).

4. CONSTRAINED DESIGN AND DESIGNED CONSTRAINTS

By conceptualizing variety available to design activity as “design space” available for exploration, Rittel uses the term constraint to describe what could be called “no-go zones” within design spaces. Thereby, the systems notion is applied to the design process as opposed to the design product.

We have thus arrived at two kinds of boundaries: One bounds elements of systems such as the parties and influences acting in a design process, or components of a design product. This kind of boundary delimits that which is considered as a system—for example, designers in a design conversation (process) or components considered for integration (product). The other bounds the degrees of freedom available for exploration. This kind of boundary delimits the conceptual territories available for design processes to unfold.

While these two kinds of boundaries may be considered as distinct, Heylighen (1995) proposes a very tight and abstract definition of the systems notion, which brings constraints, variety and systems together in a simple relationship by defining “a system as a constraint on variety” (ibid. p. 5). In this view, the two kinds of boundaries considered here appear less distinct because both can be described with this definition. Moreover, the two kinds of boundaries, even if viewed as distinct from each other, are likely to affect each other: No-go-zones in design processes contribute to choices of elements and attributes of the product, while considered product elements and attributes contribute to the establishment of no-go-zones in the process. In both cases, however, variety is reduced, i.e. possibilities eliminated.

With constraints thus impinging on designerly freedom, the question arises whether they are desirable or not, that is, whether they are beneficial or harmful. There is a far-reaching consensus amongst designers and design theorists that, perhaps counter-intuitively, reduced freedom is not necessarily harmful in design. Dorst (2003 p. 35) notes that design problems can be rather open, and that constraints can help alleviate the resulting need for “a great deal of conceptual juggling skills”. Glanville (2013) states that “[reducing variety by way of constraints] provides an advantage: many have experienced paralysis when faced with seemingly limitless variety.” This echoes what painter van Gogh wrote in a letter to his brother: “You don’t know how paralyzing it is, that stare from a blank canvas that says to the painter you can’t do anything.” (de Leeuw 1996 p. 281). Total freedom (vast variety) seems to offer unfavorable conditions for creative processes and its effects can be as
devastating as those of a total absence of freedom (no variety). Constraints (reducing variety) can also offer catalytic benefits in the creative process. Boden (1990 p. 95) notes that “far from being the antithesis of creativity, constraints on thinking are what make it possible”. Gänsfurther (2011) notes that design constraints can offer a “disciplining” (ibid. p. 157) effect, which, paradoxically, even if arbitrarily self-imposed, can have desirable effects on creativity (ibid. p. 200). Dadich (2009) confirms this from an applied perspective, explaining the use of constraints in magazine design. The ability of designers to deal with vast variety seems to come with experience. Design educators, accordingly, tend to constrain studio assignments for first-year design students extensively, and allow increasing degrees of freedom in assignments for higher years. Similarly, experienced designers can often be observed to deliberately (self-) impose constraints as a creative tool when addressing unmanageably large variety. These might be constraints of many kinds, including habits, agendas, idiosyncrasies and other ways of reducing “solution spaces”, which may also offer benefits by contributing to recognizable style and identity within bodies of creative work.

The difference between novice designers finding their assignments constrained by their educators and senior designers choosing to constrain their own work raises the question of where design constraints originate. A useful demarcation to address this question has already been introduced above with the systemic inside/outside distinction. Some designers may find it convenient or appropriate (due, say, to educational influences or professional power structures) to take constraints as naturally given by higher forces and therefore as non-negotiable. Examples include the uncritical acceptance of problem statements, user need specifications set out by marketing departments, limits of manufacturability, costing frameworks, or regulatory codes. Being in charge of distinguishing and bounding the relevant systemic elements and spaces considered in design projects, designers are responsible to at least consider re-negotiating imposed constraints in one way or another. For example, problem statements can be re-formulated, target user groups can be re-considered, manufacturing processes can be developed, price points can be challenged and cases can (hopefully) be made where the spirit of regulatory codes is obeyed while their letters are violated. Rittel explains: “[E]very constraint or limitation I pose on my action space is a decision, or at least an implicit indication of resignation” (Protzen and Farris 2010 p. 159), and he notes: “You can always seek an exception” (ibid. p. 193).

The possibility to transcend designed design spaces thus appears less as a matter of creative freedom than as a matter of ethical responsibility. Von Foerster (2003 p. 298) distinguishes between ethics “I shall...” and moral codes “thou shalt not...” based on exactly this distinction: do constraints originate within (ethics) or from outside (moral code)? The ethical implication for design and design education is obvious: Only when the freedom to make decisions is granted is there an obligation to take responsibility for one’s decisions. If this freedom is not granted, then there is no obligation to take responsibility for design choices and actions (and thus anything may be excused with the so-called Nuremberg defense: “I was just following orders!”).

5. THE ETHICS OF CONSTRAINT: THREE SCENARIOS

Despite the ethical necessity to negotiate constraints from within, the imposition (and oftentimes unquestioning acceptance) of constraints from outside is not uncommon. Three scenarios come to mind.

One is the already-mentioned educational scenario. Since the vast variety of unconstrained “design spaces” can overwhelm novice designers who have yet to learn to constrain their design space explorations by themselves, design educators impose constraints, somewhat from the outside. (There is an element of conspiracy in the relationship between design students and design tutors, the degree of which renders the question of whether educationally motivated constraints originate inside or outside of projects somewhat arguable.) This strategy should be no more than a temporary prosthesis that is used until students have gradually established habits, confidence and values that allow self-constraining and thereby making unmanageable variety more manageable.

Another one is the industrial scenario in which professional designers serve within power hierarchies. No-go zones within design spaces are dictated to designers and remain largely
non-negotiable because hierarchical organizations tend to listen upwards, not downwards (Fischer 2013) and (usually profit-oriented) design decisions taken at higher levels of authority take priority. Hopefully, the fields and corridors left open for exploration within design spaces in this way are sufficient to still allow professional designers to practice in ways that are consistent with their values and ethics. But this is not always the case.

Yet another is the computational design research scenario in which digital design process models (mechanisms) are investigated for academic research purposes such as investigations into design space modeling (for example Woodbury and Burrow 2003) or so-called multiple-constraint satisfaction (for example Emdanat 1999). Since computers cannot (in the human sense) re-consider value judgments, they cannot re-consider algorithmic no-go zones (in the designerly sense) unless instructed to do so – by humans looking in from the outside. This is easily resolved by re-considering the boundaries of the designing system in question so that the human is included: “The computer is only an arc of a larger circuit which always includes a man and an environment from which information is received and upon which efferent messages from the computer have effect” (Bateson 1972, p. 317). Viewed in this way, constraints are set by the human, from within.

6. FROM GOAL-ORIENTED TO CONSTRAINT-ORIENTED PROCESSES

With the transitions from first to second generation design methods, and from first to second order cybernetics, systems-based design theory transcended the original teleological (goal-oriented) focus of early systems and control theory towards a conversational focus. This section contrasts the teleological approach with the conversational approach.

Teleological processes are based on control and move towards defined goals. Control is achieved when those engaged in design processes (consider the straight-forward scenario of two designers) are engaged with each other as controller and controlled. The key requirement for this kind of engagement is that the variety in the controller must be identical to the variety in the controlled, so as to eliminate any ambiguity in the engagement. With defined goals, it is possible to define suitable paths towards the goals. Goal-oriented processes move along such paths, and check deviations from them along the way. These deviations are considered errors, prompting corrective measures. In a goal-oriented process, attention stays focused on the future desired state and constraints are treated as obstacles that have to be navigated while pursuing that goal: goals give direction, while constraints limit freedom.

Conversational processes, alternatively, are based on interaction and avoidance of undesirable outcomes (constraints, rather than goals). Interaction is achieved when those engaged in design processes (consider again the straight-forward scenario of two designers) are engaged with each other on an equal footing. A key feature of this kind of engagement is that it changes the variety of the conversants so as to get, at least to some extent, out of control (Glavine 2000). Focusing on avoidance (constraints) leaves open “corridors” and “fields” of possibilities within the “solution space”. Constraints, in this approach, do not necessarily dominate movement, but, as described above, may offer benefits of increased manageability and creative catalysis while still allowing some freedom for exploration. In a constraint-oriented process, attention stays focused on all perceived and possible constraints, recognizing them not as obstacles to be navigated but as enabling of ideas that would otherwise not be imagined. Outcomes remain floating (uncertain) within the space defined by the constantly changing constraints: constraints give direction, while goals limit freedom.

The question arises of how applied design processes can terminate in the absence of defined goals. How do we know when to stop? Arguably, design processes may experience difficulty stopping “from within” (there is almost always potential for incorporating additional ideas into a design). As Rittel (1973 p. 162) explains, applied design processes depend on the availability of resources (time and money, essentially), and they terminate when these resources run out. Separate from the limitation on resources, the potential contribution (variety) of new ideas in the design process decays as a design approaches a stable configuration, at which time terminating procedures may be applied to produce the finishing touches. These terminating procedures can be seen as self-imposed constraints that serve to
avoid prolonging a process in which the designers are no longer interested (their work is done), as well as to avoid running out of resources.

It is possible to engage in goal-oriented and constraint-oriented processes simultaneously (in parallel, so to speak). Since they are opposing approaches, such a dual process is best described as a dialectical one. Constraint-oriented processes are primarily concerned with what one does not want, rather than with what, specifically, one wants. They are, thus, approaches of avoidance of undesirable outcomes more than achievements of desirable outcomes.

However, any goal can be turned into a constraint, and any constraint into a goal. Avoidance of undesirable outcomes can be treated as a goal, for example. Nonetheless, there is a fundamental difference in thinking between the two approaches, with a goal-oriented approach representing a point (or vector) way of thinking in which future achievement is the driving force and a constraint-oriented approach representing a spatial (or even topological) way of thinking in which avoidance in the here and now is the driving force. The former focuses on probable outcomes, and narrowing them, the latter on possible outcomes, and expanding them. There is also a difference in the ways of dealing with time: a goal-oriented approach relying on a linear, present to future, concept of time, and a constraint-oriented approach focusing on an always changing present (multiple concepts of time).

We contend that the “desirability” of outcomes realized under a constraint-oriented approach, to the people who will engage with and be impacted by the designs implemented, can be far greater than those realized under a goal-oriented approach. Of course, this depends on how desirability is defined, and that is where social, political and economic agendas come into play. A constraint-oriented approach lends itself naturally to a true participative/democratic process (see Richards 2007).

7. CONSTRAINT REVERSAL

Constraints acting on design processes can, by way of deliberate defiance, be used as a design exploration strategy without necessarily compromising the values that motivate given sets of constraints. This may be desirable and valuable simply as a creative idea generation technique, or as a means to widen “corridors” and “fields” in constraint-oriented exploration processes if they are considered too narrow, or as a device for addressing overbearing constraints imposed from outside the design process. This move, which we refer to as “constraint reversal”, challenges the context of given design specifications and constraints, somewhat along the lines of a proposal made by Brün (1970). Based on this strategy, a statement aiming to establish a no-go zone in the “design space” is inverted, and then conditions are sought under which the inverted statement would be desirable. Consider as an example an urban planning project in which a design constraint is established with the statement: “Motor traffic must not be obstructed”. Instead of adhering to this constraint, designers might ask the following: “What would be the conditions under which motor traffic obstruction would be desirable?” This could lead to design proposals such as the establishment of a pedestrian-only area, which would in principle constitute an obstruction of motor traffic, but not necessarily be at odds with the values that motivated the original constraint against motor traffic obstruction.

8. CONCLUSION

In this paper, we first discussed the development of the changing notion of constraints in design along with the changing notion of systems since the mid-20th century in the intersection of design theory and systems theory. We showed this development starting with Rittel’s adoption of the notion of variety from Ashby, and leading towards Heylighen’s definition of “system as a constraint on variety”. Our review of design-theoretical discussions of constraints shows a far-reaching consensus that, contrary to the common way of thinking of constraints as limiting freedom for creative exploration, giving attention to constraints in design can enable the creative process. We described the goal-oriented approach in which goals give direction while constraints limit freedom in contrast to the constraint-oriented approach in which constraints give direction while goals limit freedom. A key benefit of the
constraint-oriented approach from a design perspective is that it offers “corridors” and “fields” for exploration rather than a narrow, predetermined path.

A consequence for design education of the constraint-oriented approach is the introduction of a way of thinking that is not goal-oriented, not achievement-focused, but rather one in which constraints are treated as opportunities and outcomes as secondary to the process. If everyone thought this way, we would arguably have a very different society and world than what we have now. The educational process does, however, have to recognize that we do not live in that world; and, if students are going to be able to make a living in the current world, they also have to learn how to incorporate into their designs the “goals” of those who are footing the bill. Learning how to turn the goals of these privileged individuals into desires treated as constraints would be a part of a constraint-oriented approach to design education. Therefore, in spite of the current socio-economic-political structures of modern societies, designers can still be creative and arrive at designs that point in desirable directions, even if the designs themselves are not direct forces for social change in that direction.

Finally, we introduced the strategy of constraint reversal, which allows creative exploration by means of constraint defiance without necessarily compromising the values that motivated the constraints in question in the first place. We have thus developed the role of constraints in design from non-negotiable to negotiable, and further to negotiation-provoking, implying a shift from product (oriented) design to social(ly oriented) design.

REFERENCES:


