

Collaboration and complexity: an experiment on the effect of multi-actor coupled design

Paul T. Grogan¹  · Olivier L. de Weck²

Received: 1 September 2015 / Revised: 13 January 2016 / Accepted: 18 January 2016 / Published online: 9 February 2016
© Springer-Verlag London 2016

Abstract Design of complex systems requires collaborative teams to overcome limitations of individuals; however, teamwork contributes new sources of complexity related to information exchange among members. This paper formulates a human subjects experiment to quantify the relative contribution of technical and social sources of complexity to design effort using a surrogate task based on a parameter design problem. Ten groups of 3 subjects each perform 42 design tasks with variable problem size and coupling (technical complexity) and team size (social complexity) to measure completion time (design effort). Results of a two-level regression model replicate past work to show completion time grows geometrically with problem size for highly coupled tasks. New findings show the effect of team size is independent from problem size for both coupled and uncoupled tasks considered in this study. Collaboration contributes a large fraction of total effort, and it increases with team size: about 50–60 % of time and 70–80 % of cost for pairs and 60–80 % of time and 90 % of cost for triads. Conclusions identify a role for improved design methods and tools to anticipate and overcome the high cost of collaboration.

Keywords Collaborative design · Complexity · Design theory · Parameter design · Systems engineering

This work was funded by a National Defense Science and Engineering Graduate (NDSEG) fellowship. This research was performed at Massachusetts Institute of Technology.

✉ Paul T. Grogan
pgrogan@stevens.edu

¹ Stevens Institute of Technology, Hoboken, NJ 07030, USA

² Massachusetts Institute of Technology, Cambridge, MA 02139, USA

1 Introduction

System design requires teams of people to work together to overcome individual limitations on cognition and knowledge (Arias et al. 2000). Teamwork produces multifaceted effects including benefits from parallel work flows, multiple perspectives on the problem, and specialization, but also impediments from feedback delays, misalignment of objectives, and poor group dynamics (Cohen and Bailey 1997; Kerr and Tindale 2004). While efficiency demands separation of knowledge, shared understanding improves performance in simultaneous tasks or when specialties are hard to advance, difficult to explain, or highly constrained (Postrel 2002).

Engineering organizations employ systems engineering (SE) as a structured process to achieve system requirements while distributing design work among large teams of various disciplines (Haskins 2011). SE activities iteratively flow requirements down to detailed design levels and integrate realized products up to the system level (National Aeronautical and Space Administration 2007). While model-based methods are in development (e.g., Friedenthal et al. 2012), most current SE practices use documents such as requirements and interface control specifications to exchange information across design levels.

The large scale and scope of systems projects introduce significant sources of complexity broadly interpreted as uncertainty in meeting requirements (Suh 1999). Traditional SE activities attempt to minimize social sources of complexity with strong centralized control to define interfaces and enforce trade-offs. Newer concepts such as system-of-systems allow greater degrees of operational and managerial independence among constituent systems (Maier 1998) and require a federated approach with greater degrees of autonomy (Sage and Cuppan 2001).

Collaborative design challenges traditional SE practices to structure technical design activities among a team of stakeholders or decision makers with independent or competing objectives (Lu et al. 2007). Its study goes beyond most existing engineering research to consider both technical and social features of design. As an initial step to characterize broader sources of complexity, this paper quantifies and compares the relative costs of technical and social complexity using empirical data from a human subjects experiment. Understanding the fundamental costs to collaboration will help motivate and assess new methods to accommodate distributed authority in design.

2 Background and research objectives

2.1 Complexity in design

Despite decades of SE practice and experience, large engineering projects face continued effort overruns on cost and schedule. This behavior seems to be independent of domain, time, and geography, as evidenced by numerous examples:

- 47 of 134 major US defense acquisition programs between 1997 and 2009 had a total of 74 cost breaches triggering congressional action (US Government Accountability Office 2011),
- 13 of 40 NASA Earth and space science missions between 1989 and 2010 experienced excessive cost growth with average growth exceeding 20 % (National Research Council 2010),
- A global analysis of 258 infrastructure and public works projects between 1910 and 1998 shows an average of 28 % cost growth (Flyvbjerg et al. 2002), and
- An analysis of 10 surveys on software effort estimation between 1984 and 2002 finds most projects (60–80 %) encounter effort or schedule overruns on the order of 30–40 % (Mølckken and Jørgensen 2003).

These studies attribute effort overruns to various factors including engineering and design issues, schedule issues, and quantity changes (US Government Accountability Office 2011), optimistic and unrealistic estimates, project instability, and funding issues (National Research Council 2010), and strategic misrepresentation (Flyvbjerg et al. 2002).

More broadly, project management literature attributes effort overruns to positive feedback structures in complex dynamic systems (Lyneis et al. 2001), the most dominant being rework with associated factors for productivity and work quality (Cooper et al. 2002). Conventional project management methods may trigger these feedback effects in large or interdependent, uncertain, and time-constrained

projects due to complex and counterintuitive behaviors (Williams 2005).

A unifying perspective broadly views complexity as uncertainty in meeting desired functional requirements within specified constraints (Suh 1999). More complex designs have a greater chance to exceed constraints for a fixed set of requirements compared to simpler ones, even though complexity may enable other desired features. For example, complexity from performance characteristics and airframe materials is a major source of cost escalation in fixed-wing aircraft but also enables lower weight, higher speed, and advanced capabilities (Arena et al. 2008). More general sources of complexity include structural and behavioral design features, contextual and temporal factors outside the control of designers, and perceptual factors related to stakeholder preferences and biases (Rhodes and Ross 2010).

Measurement and management of complexity may improve estimates of design effort (Bashir and Thomson 2001) and reduce overruns. A large body of literature defines and quantifies complexity in system design (e.g., Braha and Maimon 1998b; Bashir and Thomson 1999; El-Haik and Yang 1999; Ameri et al. 2008; Summers and Shah 2010). Most metrics consider the effect of structural features of system size (e.g., lines of code, number of components, number of functions) or degree of coupling on design cost. Fewer studies investigate the effects of complexity on individual performance, although it is generally perceived to contribute more errors and lower productivity (Card and Agresti 1988).

Several human subjects experiments empirically study complexity in design. Hirschi and Frey (2002), hereafter referred to as H&F, quantify the effect of problem size and coupling on task completion time using a surrogate parameter design task. Normalized completion time grows linearly with the number of uncoupled variables but geometrically with coupled variables—much faster than polynomial growth in numerical solvers. Differences may be explained from a cognitive psychology perspective of limited short-term memory (Miller 1956). Another study assembling molecular models as a surrogate design task also finds effort grows super-linearly with a structural complexity metric via a power law with exponent 1.48 (Sinha 2014). A third study using well-defined building design problems and fixed allotted effort finds increasing problem scale exponentially decreases solution quality (Flager et al. 2014). These studies consistently show technical complexity increases effort to meet fixed requirements or decreases quality under fixed effort.

2.2 Collaboration in design

In contrast to an individual, a design team has no single memory and requires communication to exchange

information and construct knowledge among members (Konda et al. 1992; Arias et al. 2000). External artifacts such as models, documents, or tools extend natural limits of memory and communication. They often take advantage of computational information systems (Engelbart 1995); however, there exist benefits to physical media as well (Arias et al. 1997). General group performance is not a simple function of its individuals and is instead correlated with attributes such as social sensitivity and equality in distribution of conversation turn-taking (Woolley et al. 2010).

Most literature on teamwork or group performance exists in the fields of social psychology (e.g., Kerr and Tindale 2004) and organizational or management science (e.g., Cohen and Bailey 1997). Negative effects of group size described as social loafing or the Ringelmann effect are attributed to coordination and motivation losses (Kravitz and Martin 1986; Ingham et al. 1974). Other factors impacting group performance include cohesion composed largely of group pride (Mullen and Copper 1994), friendship mediated through cooperation and commitment (Jehn and Pradhan 1997), task and team familiarity (Goodman and Leyden 1991), and trust mediated through motivation (Dirks 1999). Clustered organizational structures made up of cliques enable higher group performance (Huberman and Hogg 1995; Kearns et al. 2006; McCubbins et al. 2009); however, clustering can also restrict exploration in design (Lazer and Friedman 2007; Mason and Watts 2012; Shore et al. 2015). While relevant for understanding how team composition and structure affects performance, these studies do not simultaneously address the contextual effects of design tasks.

Two studies address the effect of component complexity, defined as the number of unique actions required to complete a task (Wood 1986), on group performance. Weingart (1992) finds component complexity increases both the amount and quality of planning for some aspects of a task and decreases group effort, both effects mediating lower group performance. Argote et al. (1995) find component complexity has a negative main effect on group performance. Performance gains for simple products are greater than complex products as groups gain experience. Turnover of group members also has a larger effect on simple tasks compared to complex ones, possibly due to social loafing behaviors. However, these studies only consider well-defined and prescribed tasks such as assembling origami and craft structures rather than the more creative process of design.

Collaborative engineering applies social science research to improve design outcomes among a team of stakeholders with a common goal but limited resources or conflicting interests. Lu et al. (2007) frame engineering collaboration as a negotiation with four steps: (1)

interaction among designers to (2) construct a common understanding leading to (3) a group preference, and finally to (4) attain agreement on a design. Research from organizational science, social psychology, social choice, and decision sciences apply to each step (respectively) to develop new design approaches. Other literature presents methods to improve collaborative design such as repeatable processes to be conducted by practitioners (Briggs et al. 2003) and software tools to improve information exchange among designers (Wang et al. 2002).

2.3 Research objectives

Literature in engineering design emphasizes technical complexity with limited consideration of features relevant for multi-actor design. Social science research investigates factors contributing to group performance without considering the unstructured technical activities in engineering design. Collaborative design research crosses both domains; however, most literature emphasizes interventions to improve outcomes instead of more basic processes of design. In particular, no existing study quantifies the effect of technical and social complexity on design effort. To address the intersection between these topics, this paper asks: *What are the relative costs of technical and social complexity in multi-actor design under barriers to collaboration?*

Past work shows technical complexity from the design task super-linearly increases effort for individuals and similar results are expected for groups. Social complexity from the design team likely also increases effort, but it is not known by how much or if it interacts with technical complexity. Collaborative design relies on information exchange to construct shared knowledge. Its cost depends on the efficiency and effectiveness of communication which is inherently limited by cognition, language, and organizational boundaries. These barriers may be simulated in artificial design activities by purposefully limiting communication among designers as a bounding case on realistic design activities.

To assess the research objectives, this study performs a human subjects experiment using surrogate design tasks with purposeful barriers to collaboration to limit and control communication among designers. Tasks vary elements of technical complexity (problem size) and social complexity (team size) to measure required completion time (design effort). Although limited due to simplifications of surrogate tasks performed in a controlled environment, results are expected to bound realistic design tasks and provide a measure of the relative contributions of technical and social complexity to design effort. Improved understanding of the technical and social costs to collaboration will improve estimates of required effort and lead to future

work to assess new methods and tools to improve collaborative design.

3 Multi-actor system design model

This section develops a parameter-based model of multi-actor system design as a surrogate task. It provides experimental control over technical complexity by varying the number and degree of coupling between variables and social complexity by assigning variables to multiple designers. The surrogate task also removes all context from the design problem to reduce effects of domain knowledge or experience and allow tasks to be solved in a short time period.

3.1 Single-actor surrogate design task

Parameter design maps functional requirements (FRs) to design parameters (DPs) (Suh 1999). In its most general form, a system model M in Eq. 1 transforms an input vector \mathbf{x} of DPs to an output vector \mathbf{y} of functional characteristics.

$$M(\mathbf{x}) = \mathbf{y} \quad (1)$$

An error function E in Eq. 2 evaluates FR constraints by comparing design outputs \mathbf{y} with requirements \mathbf{y}^* . A design meets FRs if error does not exceed a tolerance \mathbf{E}^* .

$$E(\mathbf{y}, \mathbf{y}^*) \leq \mathbf{E}^* \quad (2)$$

Combining Eqs. 1–3 states the objective of a parameter design task.

$$\text{find } \mathbf{x} \text{ s.t. } E(M(\mathbf{x}), \mathbf{y}^*) \leq \mathbf{E}^* \quad (3)$$

Past work in H&F assumes a particular form for M , E , and \mathbf{E}^* . A linear system model in Eq. 4 relates N inputs and M outputs with an $M \times N$ transformation matrix \mathbf{M} . Element m_{ij} quantifies coupling between output i and input j , i.e., $m_{ij} = dy_i/dx_j$ for a linear system.

$$M(\mathbf{x}) = \mathbf{M}\mathbf{x} = \begin{bmatrix} m_{11} & \dots & m_{1N} \\ \vdots & \ddots & \vdots \\ m_{M1} & \dots & m_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} \quad (4)$$

An error function in Eq. 5 takes the absolute value of the difference between design outputs and FRs.

$$E(\mathbf{y}, \mathbf{y}^*) = \{|y_i - y_i^*|\} \forall i \quad (5)$$

By substituting Eqs. 4, 5 in Eq. 3 and assigning a fixed error tolerance $E_i^* = \varepsilon$ for all FRs, Eq. 6 states the objective for a single-actor surrogate design task.

$$\text{find } \mathbf{x} \text{ s.t. } \left| \sum_j m_{ij} x_j - y_i^* \right| \leq \varepsilon \forall i \quad (6)$$

The variant with $N = M$ and $\varepsilon = 0$ is a linear system of equations which could be solved with Gaussian elimination in approximately $2N^3/3$ operations (Gentle 1998) if all required parameters are quantified by a central actor.

3.2 Multi-actor surrogate design task

The multi-actor surrogate design task extends the single-actor case by assigning each input and output to one designer to represent control over DPs and FRs. Assignments for a design task with n designers are formalized by two binary (0, 1) matrices. An $n \times N$ matrix \mathbf{I} assigns inputs where element I_{ij} is defined in Eq. 7.

$$I_{ij} = \begin{cases} 1 & \text{if input } j \text{ assigned to designer } i \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Similarly, an $n \times M$ matrix \mathbf{O} assigns outputs where element O_{ij} is defined in Eq. 8.

$$O_{ij} = \begin{cases} 1 & \text{if output } j \text{ assigned to designer } i \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Assignment allows social coupling to emerge when one designer's inputs affect another's outputs. An $n \times n$ square matrix \mathbf{D} captures social couplings by composing \mathbf{O} , \mathbf{M} , and \mathbf{I} in Eq. 9.

$$\mathbf{O} \times \mathbf{M} \times \mathbf{I}^T = \mathbf{D} = \begin{bmatrix} d_{11} & \dots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \dots & d_{nn} \end{bmatrix} \quad (9)$$

Element $d_{ij} \neq 0$ indicates social coupling between designers i and j , specifically that designer i 's outputs depend on designer j 's inputs.

Consider an example design task with four inputs, four outputs, and three designers in Eq. 10. Designer A controls inputs 1 and 2 and outputs 1 and 2. Designer B controls input 3 and output 3. Designer C controls input 4 and output 4.

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & 0 \\ m_{21} & m_{22} & m_{23} & 0 \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & m_{44} \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{O} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

The resulting \mathbf{D} matrix in Eq. 11 shows coupling between designers. Designer A's inputs affect outputs for A and B, B's inputs affect all three designers' outputs, and C's inputs only affect its own outputs.

$$\mathbf{D} = \begin{bmatrix} (m_{11} + m_{12} + m_{21} + m_{22}) & (m_{13} + m_{23}) & 0 \\ (m_{31} + m_{32}) & m_{33} & m_{34} \\ 0 & 0 & m_{44} \end{bmatrix} \quad (11)$$

Figure 1 illustrates a two-layer hyper-graph of the design task. The technical layer shows coupling between inputs and outputs from \mathbf{M} . The social layer shows coupling between designers from \mathbf{D} . Input and output assignments derived from \mathbf{I} and \mathbf{O} , respectively, connect the two layers.

3.3 Task generation method

A method adapted from H&F defines \mathbf{M} and \mathbf{y}^* for randomized tasks. The input range is bounded by $x_i \in [-1, 1]$ and the initial value of all inputs and outputs is zero ($\mathbf{x}_0 = \mathbf{y}_0 = 0$). Coupled tasks generate \mathbf{M} with $m_{ij} \in (-1, 1)$ by composing orthonormal bases of random vectors drawn from a uniform (0, 1) distribution. The resulting orthonormal matrix guarantees well-conditioned and balanced relationships between inputs and outputs and consistency of a single solution. Uncoupled tasks generate diagonal \mathbf{M} with elements $m_{ii} \in \{-1, 1\}$ each selected with probability 0.5.

A target output \mathbf{y}^* with $y_i^* \in (-1, 1)$ is the orthonormal basis of a random vector drawn from a uniform (0, 1) distribution subject to a constraint that each input of the solution $\mathbf{x}^* = \mathbf{M}^{-1}\mathbf{y}^*$ must be greater than $\delta = 0.05$ units from the initial conditions, i.e., $|x_i^* - x_{0,i}| > \delta \forall i$. The resulting target has a Euclidean norm of 1 (i.e., $\|\mathbf{y}^*\| = 1$) to provide a standard solution distance and requires a

minimum change in δ for each input to achieve the zero-error solution.

3.4 Software implementation

A distributed software application implements the surrogate design task with three purposeful barriers to cognition and collaboration. First, the system model is hidden so designers can only observe effects of input value changes on their assigned outputs. Second, no quantitative information is displayed to prevent designers from mathematically solving the linear system. Finally, designers cannot share graphical displays to simulate communication barriers across organizational boundaries.

Figure 2 shows a designer client for a task with two vertical slider inputs and two horizontal slider outputs. Randomized labels are assigned to tasks (e.g., absorbed copper), inputs (e.g., diameter, flexibility), and outputs (e.g., epsilon, rho). To modify inputs, users drag the slider thumb up and down, press up and down keys to move 1/200 of the range or 0.01 units, or press page up and page down keys to move 1/20 of the range or 0.1 units. While dragging the slider thumb, inputs only update once released. The signal icon displays a green check mark if an output is within the error tolerance ε of the target, visually a darker region; otherwise, it displays a red cross. A designer has no controls if not assigned any inputs or outputs for a particular task.

The user interface differs slightly from H&F. First, this design displays inputs vertically and outputs horizontally, rather than both vertically, to prevent the uncoupled tasks from being solved by aligning input slider thumbs with

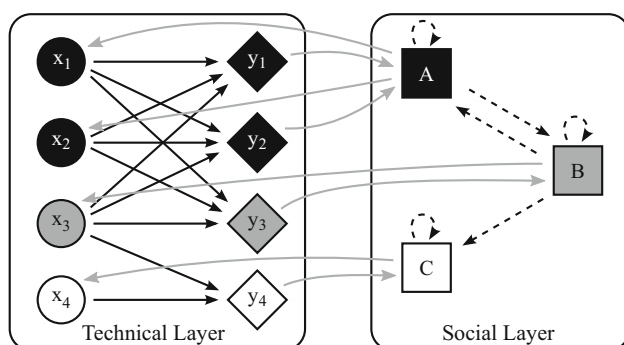


Fig. 1 Multi-actor design task can be represented as a hyper-graph connecting the inputs (x_i) and outputs (y_i) in a technical layer and designers (A: black, B: gray, C: white) in a social layer

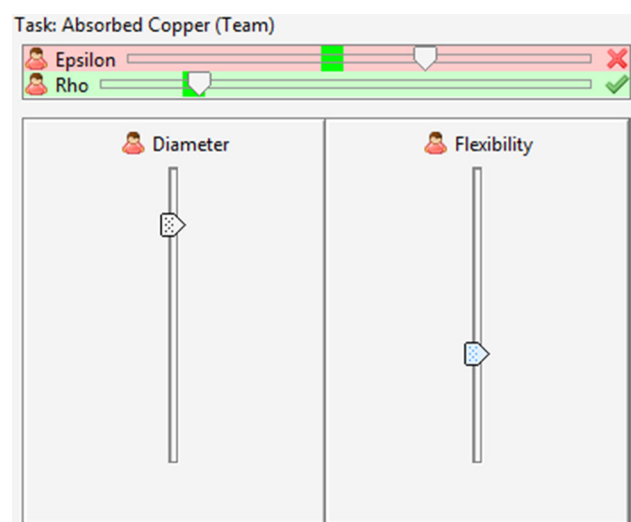


Fig. 2 Designer application includes sliders for assigned inputs and outputs. Vertical input sliders respond to user actions. Horizontal output sliders show the target and range of acceptable values

output target values. Second, this design does not require a designer to press a “Refresh Plot” button to update outputs. Instead, discrete input methods and verbal communication delays limit feedback rates with a smaller timescale compared to the previous study. Finally, the addition of randomized labels and icons and a completion audio queue improve aesthetics without an expected impact on results.

3.5 Model assumptions and limitations

The multi-actor surrogate design task makes several simplifying assumptions which limit its generalizability to broader design tasks. First, it assumes there is exactly one zero-error solution to the design task as one of three possible cases:

1. Overdetermined task with no solutions,
2. Underdetermined task with a plurality of solutions, and
3. Uniquely determined task with exactly one solution.

Overdetermined tasks have no zero-error solutions, i.e., no design meets all requirements. Either the task itself is infeasible or requirements must be relaxed to find a feasible solution. The process of relaxing requirements, operationalized as changing target outputs, ultimately produces either an under- or uniquely determined task.

Underdetermined tasks have more than one solution meeting all requirements. Preference for solutions considers other objectives such as minimizing cost or maximizing value. Without considering these objectives in the task formulation, all feasible solutions are functionally similar to a uniquely defined solution. Providing an objective function may confound experimental variables by also measuring an individual's effort to maximize value.

Uniquely determined tasks are least similar to real-world design tasks but most applicable to the experimental framing. A single-solution criterion provides a concise measurement of effort and aligns all designers' goals such that secondary objectives are not needed. While there exists one zero-error solution, the error bounds ε in this formulation provide a small range of acceptable solutions. This practical, rather than theoretical, consideration allows designers to find a solution with discrete inputs.

A second limitation arises from the linear system defined by \mathbf{M} . While most real-world systems are not linear, they also provide context such as physical laws and mathematical models encapsulated in domain knowledge. In this context-free case, even linear systems are not perceived as simple due to limited cognitive abilities without quantitative aids. Furthermore, a linear system model provides three practical advantages. First, it is the simplest model of technical coupling with minimal assumptions. Second, consistent linear systems with a single zero-error solution can be generated for arbitrarily complex design

tasks using randomized orthonormal matrices. Finally, outputs can be rapidly computed with matrix multiplication.

4 Experimental methodology

The multi-actor surrogate design formulation leads to four types of tasks:

- I. Uncoupled decisions within designers (\mathbf{M} and \mathbf{D} diagonal),
- II. Coupled decisions within designers (\mathbf{M} unconstrained, \mathbf{D} diagonal),
- III. Uncoupled decisions across designers (\mathbf{M} diagonal, \mathbf{D} unconstrained), and
- IV. Coupled decisions across designers (\mathbf{M} and \mathbf{D} unconstrained);

where the first two items are the cases studied by H&F. This study adds social coupling to comparatively evaluate type III and IV design tasks.

4.1 Experimental design

The experiment is structured as a multi-level study with tasks and design teams as hierarchical units of analysis. The number of task variables N and degree of coupling operationalize technical complexity. The number of coupled designers n operationalizes social complexity. The time t to complete a task operationalizes design effort.

The experimental design in Table 1 samples tasks for $1 \leq n \leq 3$ designers and $N = M$ inputs and outputs with $2 \leq N_C \leq 4$ coupled and $2 \leq N_U \leq 6$ uncoupled variables. The 14 numbered design tasks address the following objectives:

1. Validate previous results: five cases (tasks 1, 2, 4, 5, 6).
2. Vary social coupling while holding technical coupling constant: three cases with three levels each (tasks 1, 7, 11; 5, 9, 13; 6, 10, 14).

Table 1 Design task type with n coupled designers and N input and output variables

n	N_U (uncoupled)					N_C (coupled)		
	2	3	4	5	6	2	3	4
1	I ^a	I ₁ ^a	I ₂ ^a	I ^a	I ₃	II ₄ ^a	II ₅ ^a	II ₆ ^a
2	III	III ₇	III	III	III	IV ₈	IV ₉	IV ₁₀
3	–	III ₁₁	III	III	III ₁₂	–	IV ₁₃	IV ₁₄

Task number in this experimental design

^a Task in H&F

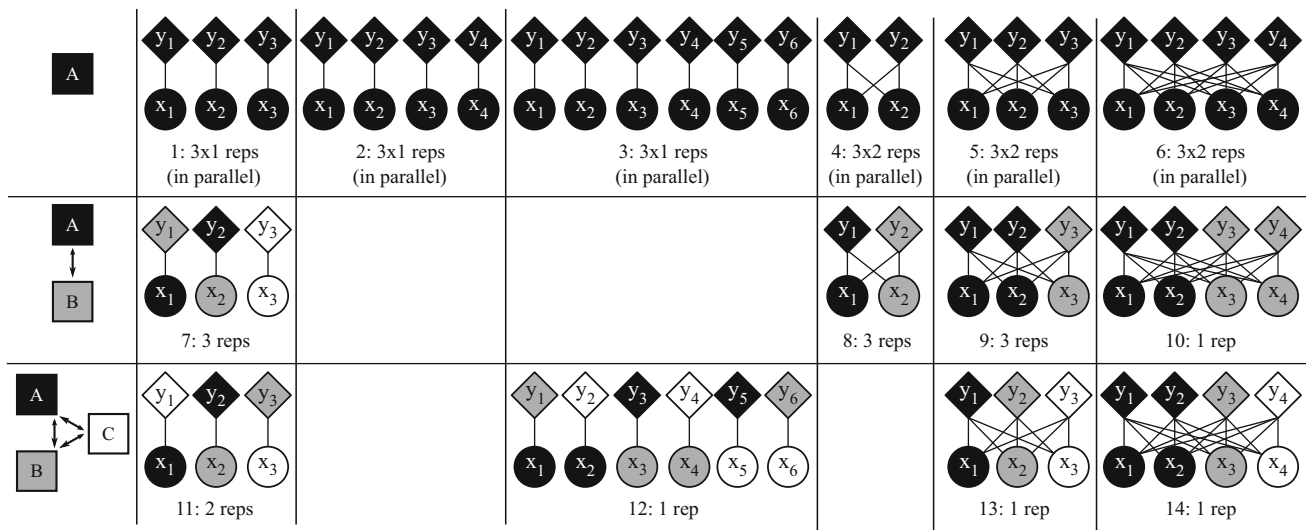


Fig. 3 Graph representation of the six uncoupled tasks and eight coupled tasks with input/output assignments identified by *color*

3. Vary technical coupling while holding nonzero social coupling constant: one case with three levels (tasks 8, 9, 10).

Figure 3 illustrates designer assignments and replications for the 14 task types. Individual tasks are conducted in parallel for the three designers, and pair tasks use rotating assignments for each replication. In total, this experimental design calls for 42 tasks: 3×9 individual and 15 team.

4.2 Subjects

Ten sessions of three subjects participated in this study. Volunteers were recruited from email solicitation and a convenience sample of graduate engineering programs at MIT and were not paid. Table 2 summarizes complete subject demographics. Subjects were predominately male (66.7 %) and 25–29 years of age (60.0 %). Most subjects had never interacted with each other in the past (50.0 % of pairs).

Table 2 Experimental subject demographics in six sessions

Category	Value	Count	(%)
Gender	Male	20	66.7
	Female	10	33.3
Age	18–24	9	30.0
	25–29	18	60.0
	30–34	2	6.7
	35–39	1	3.3
Frequency of past interactions with other subjects	Never or nearly never	30	50.0
	Occasional (monthly)	22	36.7
	Frequent (weekly)	8	13.3

4.3 Experimental procedure

Design sessions are scheduled based on the availability of three volunteers to form ad hoc teams. There is neither random assignment of subjects to sessions nor purposeful selection. All experiments are conducted in university classrooms using a standard layout. Subjects choose color (red, green, or blue) and sit on one side of a four-seat rectangular designer table with the fourth seat reserved for the administrator. The table is arranged such that each computer display is only visible to the seated individual.

Experimental sessions are conducted using an IRB-approved protocol. Participants may exit the study at any point but no such events occurred. A scripted presentation introduces the experimental objectives and issues consent forms and a demographic questionnaire. A series of five training tasks introduce subjects to the software and design process. Training tasks are similar to experimental tasks and increase in difficulty from an $n = 1$, $N_U = 2$ to an $n = 3$, $N_C = 3$ task. During training, an administrator explains the software interface, design objectives, and communication limitations. Training takes approximately 15 min to complete.

Four sessions complete the design tasks in randomized order. Six other sessions use a partially randomized order with a constraint that no $N_C = 4$ design problems (tasks 6, 10, and 14) can occur within the first ten tasks. This constraint acknowledges learning effects to avoid subjects feeling overwhelmed by large design problems early in the session. The order of each task is recorded to control for learning effects in analysis. There is no time limit on solving each task, although participants are instructed the expected time to complete all tasks is 60 min. All designer input modifications are automatically logged to file and the

administrator advances to the next task when all subjects are ready.

5 Results and analysis

5.1 Experimental results

Table 3 summarizes raw experimental results. Several tasks omit missing data (NAs) from administrative and technical issues. One replication of tasks 4 and 5 is missing from three sessions, and one replication of task 6 is missing from six sessions due to a modified experimental design. The reduced amount of data in these sessions is not expected to bias overall results. Furthermore, two sessions use $\varepsilon = 0.1$ and one uses $\varepsilon = 0.11$ while the remaining seven use $\varepsilon = 0.05$.

Two other issues result in missing data. One individual replication of task 6 was removed due to subject concession which biases results due to subject mortality effects; however, this is the only observed instance in all tasks (1/30 incidence). One replication of task 3 was removed due to technical problems associated with the network connection.

5.2 Analysis method

H&F use the linear regression models in Eqs. 12, 13 to describe the effect of uncoupled task size N_U and coupled task size N_C on completion time \bar{t}_i for task i normalized by a $N_C = 2$ task completed at some point during the session.

$$\bar{t}_i = \beta_0 + \beta_1(N_U) + r_i \quad (12)$$

$$\log(\bar{t}_i) = \beta_0 + \beta_1(N_C) + r_i \quad (13)$$

Task time normalization imperfectly addresses correlated samples within subjects which otherwise violates the independence of errors required by linear regression. Following recommendations of Flager et al. (2014), this study uses a conditional two-level regression model (Raudenbush and Bryk 2002) also known as a hierarchical or mixed-effects model with general form in Eq. 14.

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \sum_i \beta_{ij}(X_i) + r_{ij} \quad (14a)$$

$$\text{Level 2: } \beta_{ij} = \gamma_{i0} + \sum_j \gamma_{ij}(Z_j) + u_{ij} \quad (14b)$$

A two-level regression accommodates correlated samples at level 1 by allowing coefficients to vary with a level 2 factor. In this study, level 1 predictors (X_i) relate to task i and level 2 predictors (Z_j) relate to group j where individuals and teams are considered separate groups. Level 2 equations allow the intercept (β_{0j}) or slopes ($\beta_{ij}, i \neq 0$) to vary based on group-specific factors. Errors exist at both the task (r_{ij}) and the group (u_{ij}) levels. This study uses the lme4 R package for linear mixed-effects models for analysis (Bates et al. 2015).

Plausible level 1 predictors include problem size N_U or N_C , team size n , error tolerance ε , and task order O . Based on H&F, problem size is expected to have a geometric impact on task time $\log(t) \propto N$ to model a random sampling of a N -dimensional space. Team size is hypothesized to have a power law relationship with task time

Table 3 Summary of raw experimental results

Task	n	N	C/U	Samples			Task completion time (s)				
				Num.	NAs	Net	Min.	1st Q.	Med.	3rd Q.	Max.
1	1	3	U	30	0	0	2.8	10.0	13.7	16.1	33.5
2	1	4	U	30	0	0	3.8	11.1	14.9	19.2	51.2
3	1	6	U	30	3	27	9.3	22.9	27.6	44.4	56.9
4	1	2	C	60	9	51	3.7	9.1	13.8	20.4	40.0
5	1	3	C	60	9	51	4.0	29.3	36.8	57.8	201.9
6	1	4	C	60	19	41	18.3	59.1	105.9	148.1	368.0
7	2	3	U	30	0	30	10.4	22.7	32.3	37.6	61.2
8	2	2	C	30	0	30	6.6	18.9	30.6	50.0	203.0
9	2	3	C	30	6	24	27.2	89.0	127.8	228.7	476.7
10	2	4	C	10	0	10	82.9	141.5	341.0	416.4	770.8
11	3	3	U	20	0	20	17.9	29.1	33.4	43.1	89.4
12	3	6	U	10	0	10	37.5	47.8	84.8	123.1	177.8
13	3	3	C	10	0	10	86.1	97.0	132.1	326.3	783.3
14	3	4	C	10	0	10	129.9	361.4	652.8	704.7	1015.0
Total				420	46	374					

$\log(t) \propto \log(n)$. Variations in error tolerance ε are treated as independent factors. Task order quantifies learning effects accumulated in sequential task ordering O with a power law relationship $\log(t) \propto \log(O)$ based on Henderson's Law for learning curves (Henderson 1968). This analysis uses a logarithmic transform of task time ($Y_{ij} = \log(t_{ij})$) for both coupled and uncoupled tasks to accommodate expected multiplicative effects of level 1 predictors.

Plausible level 2 predictors include individual and team demographic factors for gender, age, and frequency of past interactions. Preliminary analysis finds a group identifier factor G addresses more variance than other factors. As the role of demographic factors in team composition is not the focus of this study, G is used as the only level 2 predictor similar, through much more limited in scope, to a collective intelligence factor (Woolley et al. 2010).

5.3 Uncoupled task analysis

A backward stepwise regression procedure in Table 4 sequentially eliminates nonsignificant factors. Each step reduces AIC computed using a maximum likelihood (ML) criterion. The step 1 model includes fixed effects for problem size N_U , team size $\log(n)$, task order $\log(O)$, error tolerance factor ε , and interaction between problem size and team size $N_U \times \log(n)$ and random effects of G on the intercept and problem size. The step 2 model eliminates the random effect of G on N_U as it contributes little variance. The step 3 model eliminates the interaction term $N_U \times \log(n)$ with $t(98.6) = -0.486$, $p = 0.63$ using the Kenward-Roger (KR) approximation for degrees of freedom (Halekoh and Højsgaard 2014). Graphical inspection of model residuals in a Q–Q plot indicate a nearly normal distribution.

Equation 15 shows the uncoupled task step 3 model where γ_{01} is unique to each group.

$$\text{Level 1: } \log t_{ij} = \beta_{0j} + \beta_{1j}(N_U) + \beta_{2j}(\log n) + \beta_{3j}(\log O) + \beta_{4j}(\varepsilon_{0.1}) + \beta_{5j}(\varepsilon_{0.11}) + r_{ij} \quad (15a)$$

$$\text{Level 2: } \beta_{ij} = \begin{cases} \gamma_{00} + \gamma_{01}(G) + u_{0j}, & \text{if } i = 0. \\ \gamma_{i0} + u_{ij}, & \text{otherwise.} \end{cases} \quad (15b)$$

5.4 Coupled task analysis

A backward stepwise regression procedure in Table 5 sequentially eliminates nonsignificant factors. The step 1 model includes fixed effects for problem size N_C , team size $\log(n)$, task order $\log(O)$, error tolerance factor ε , and interaction between problem size and team size $N_C \times \log(n)$ and random effects of G on the intercept and problem size. The Step 2 model eliminates the random effect of G on the intercept as it contributes minimal variance. The step 3 model eliminates the interaction term $N_C \times \log(n)$ with $t(233.8) = 0.630$, $p = 0.53$ using the KR approximation. Graphical inspection of model residuals in a Q–Q plot show only a minor deviation from a normal distribution on the positive extrema.

Equation 16 shows the coupled task step 3 model where γ_{11} is unique to each group.

$$\text{Level 1: } \log t_{ij} = \beta_{0j} + \beta_{1j}(N_C) + \beta_{2j}(\log n) + \beta_{3j}(\log O) + \beta_{4j}(\varepsilon_{0.1}) + \beta_{5j}(\varepsilon_{0.11}) + r_{ij} \quad (16a)$$

Table 4 Stepwise linear multiple effects models for uncoupled tasks

Random	Coef.	Step 1 model (AIC = 117.39)			Step 2 model (AIC = 115.39)			Step 3 model (AIC = 113.62)		
		Factor	Variance	Std. dev.	Factor	Variance	Std. dev.	Factor	Variance	Std. dev.
G	γ_{01}	(Inter.)	0.084	0.290	(Inter.)	0.084	0.290	(Inter.)	0.085	0.291
G	γ_{11}	N_U	1.6×10^{-15}	4.0×10^{-8}	N_U	–	–	N_U	–	–
Residual			0.073	0.271		0.073	0.271		0.073	0.271
Fixed	Coef.	Estimate	S.E.	t stat.	Estimate	S.E.	t stat.	Estimate	S.E.	t stat.
(Intercept)	γ_{00}	2.321	0.143	16.210	2.321	0.143	16.210	2.357	0.122	19.249
N_U	γ_{10}	0.276	0.024	11.618	0.276	0.024	11.618	0.269	0.019	14.256
$\log n$	γ_{20}	0.993	0.193	5.157	0.993	0.193	5.157	0.914	0.103	8.832
$\log O$	γ_{30}	–0.195	0.031	–6.257	–0.195	0.031	–6.257	–0.197	0.031	–6.397
$\varepsilon_{0.1}$	γ_{40}	–0.465	0.130	–3.569	–0.465	0.130	–3.569	–0.464	0.130	–3.559
$\varepsilon_{0.11}$	γ_{50}	–0.953	0.174	–5.488	–0.953	0.174	–5.488	–0.952	0.174	–5.474
$N_U \times \log n$	γ_{60}	–0.019	0.038	–0.486	–0.019	0.038	–0.486	–	–	–

Table 5 Stepwise linear multiple effects models for coupled tasks

Random	Coef.	Step 1 Model (AIC = 463.44)			Step 2 Model (AIC = 461.44)			Step 3 Model (AIC = 459.84)		
		Factor	Variance	Std. dev.	Factor	Variance	Std. dev.	Factor	Variance	Std. dev.
G	γ_{01}	(Inter.)	$3.2 \cdot 10^{-16}$	$1.8 \cdot 10^{-8}$	(Inter.)	—	—	(Inter.)	—	—
G	γ_{11}	N_C	$6.8 \cdot 10^{-3}$	0.083	N_C	$6.8 \cdot 10^{-3}$	0.083	N_C	$6.8 \cdot 10^{-3}$	0.082
Residual			0.369	0.607		0.368	0.606		0.369	0.607
Fixed	Coef.	Estimate	S.E.	t stat.	Estimate	S.E.	t stat.	Estimate	S.E.	t stat.
(Intercept)	γ_{00}	1.094	0.208	5.260	1.094	0.208	5.260	1.028	0.180	5.712
N_C	γ_{10}	1.025	0.066	15.419	1.025	0.066	15.419	1.049	0.054	19.277
$\log n$	γ_{20}	1.197	0.428	2.795	1.197	0.428	2.795	1.451	0.144	10.056
$\log O$	γ_{30}	−0.207	0.051	−4.085	−0.207	0.051	−4.085	−0.209	0.051	−4.126
$\varepsilon_{0.1}$	γ_{40}	−0.266	0.153	−1.735	−0.266	0.153	−1.735	−0.270	0.153	−1.763
$\varepsilon_{0.11}$	γ_{50}	−0.755	0.206	−3.665	−0.755	0.206	−3.665	−0.759	0.206	−3.684
$N_C \times \log n$	γ_{60}	0.086	0.136	0.630	0.086	0.136	0.630	—	—	—

$$\text{Level 2: } \beta_{ij} = \begin{cases} \gamma_{00} + u_{0j}, & \text{if } i = 0. \\ \gamma_{10} + \gamma_{11}(G) + u_{1j}, & \text{if } i = 1. \\ \gamma_{i0} + u_{ij}, & \text{otherwise.} \end{cases} \quad (16b)$$

6 Discussion

6.1 Interpretation of results

Consider the rearranged model in Eq. 17 with separate coefficients for uncoupled (B_{Ui}) and coupled (B_{Ci}) tasks.

$$t = B_0 \cdot B_1^N \cdot n^{B_2} \cdot O^{B_3} \cdot B_4^{\varepsilon_{0.1}} \cdot B_5^{\varepsilon_{0.11}} \quad (17)$$

where $B_0 = \exp(\gamma_{00})$, $B_1 = \exp(\gamma_{10})$, $B_2 = \gamma_{20}$, $B_3 = \gamma_{30}$, $B_4 = \exp(\gamma_{40})$, and $B_5 = \exp(\gamma_{50})$.

This form yields models for completion time of uncoupled tasks in Eq. 18 and coupled tasks in Eq. 19.

$$t_U = 10.6 \cdot 1.31^{N_U} \cdot n^{0.91} \cdot O^{-0.20} \cdot 0.63^{\varepsilon_{0.1}} \cdot 0.39^{\varepsilon_{0.11}} \quad (18)$$

$$t_C = 2.80 \cdot 2.85^{N_C} \cdot n^{1.45} \cdot O^{-0.21} \cdot 0.76^{\varepsilon_{0.1}} \cdot 0.47^{\varepsilon_{0.11}} \quad (19)$$

Figures 4 and 5 plot completion time contours for a task with variable problem and team size, ordered $O = 10$, and with error tolerance $\varepsilon = 0.05$. Note the large differences in completion time magnitude and contour shapes for uncoupled and coupled tasks.

Intercept coefficients with estimated values $B_{U0} = 10.6$ and $B_{C0} = 2.80$ are interpreted as the base time to consider a design problem with no prior experience. The higher uncoupled task coefficient likely captures differing impacts

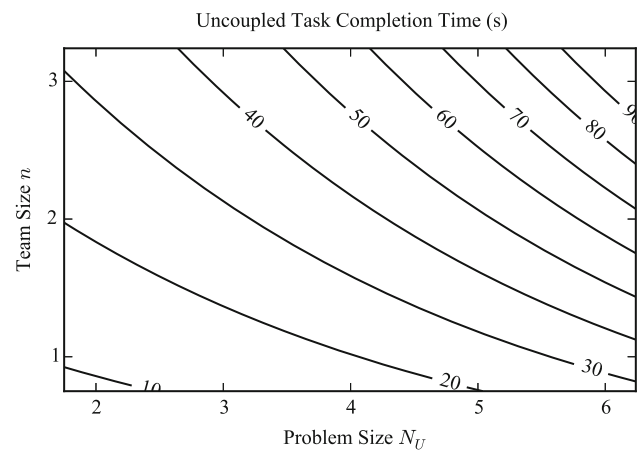


Fig. 4 Contour plot of uncoupled task completion time (seconds) for the mean design group with $O = 10$ and $\varepsilon = 0.05$

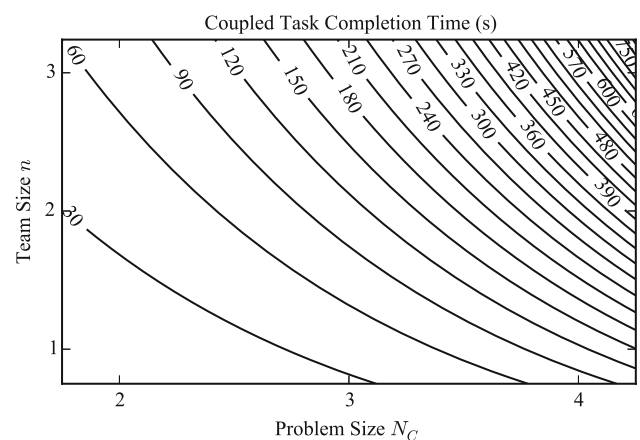


Fig. 5 Contour plot of coupled task completion time (seconds) for the mean design group with $O = 10$ and $\varepsilon = 0.05$

of error tolerance factors rather than fundamental differences in problem framing and setup compared to coupled tasks.

Results show substantially different problem size coefficients for uncoupled tasks $B_{U1} = 1.31$ and coupled tasks $B_{C1} = 2.85$. These factors indicate each additional uncoupled variable demands 31 % more time compared to 185 % for each additional coupled variable. Coupling forces designers to iterate on solutions with cognitive or communication delays (Smith and Eppinger 1997). Results are consistent with the churn effect where coupling and interdependency increase design time (Yassine et al. 2003). Although constrained to a geometric form, the effect of problem size on uncoupled tasks is largely linear similar to results of H&F. For coupled tasks, a 95 % confidence interval of [2.56, 3.18] for B_{C1} does not include the value of 3.4 previously found by H&F. This difference may be attributed to known contextual differences between this study and H&F. For example, the new user interface may provide higher information update frequency to benefit integrated design (Yassine et al. 2013).

Results also show different team size coefficients for uncoupled tasks $B_{U2} = 0.91$ and coupled tasks $B_{C2} = 1.45$. These factors indicate uncoupled tasks scale sub-linearly with team size and coupled tasks scale super-linearly. For comparison, a value of 1.0 suggests cyclic or centralized interaction (most agents interact with only a few others) and a value of 2.0 suggests complete interaction (most agents interact with most others). This polynomial factor relates to the information content of the organizational structure (i.e., **D** matrix) as theorized and demonstrated in prior work on complexity metrics (Braha and Maimon 1998a, b). Results suggest uncoupled tasks allow cyclic or centralized interaction while coupled tasks require more, but not complete, interaction between designer pairs.

The experience curve coefficients $B_{U3} = -0.20$ for uncoupled tasks and $B_{C3} = -0.21$ for coupled tasks show the time to complete the last task ($O = 24$) requires only about 50 % of the first ($O = 1$) and about 80 % of the tenth ($O = 10$). Results emphasize the importance of controlling for learning effects in analysis.

Error tolerance dummy variable coefficients are $B_{U4} = 0.63$ and $B_{U5} = 0.39$ for uncoupled tasks and $B_{C4} = 0.76$ and $B_{C5} = 0.47$ for coupled tasks. These factors are interpreted as completion time modifiers due to alternate error tolerances ($\varepsilon = 0.1$ and 0.11, respectively) compared to standard conditions with $\varepsilon = 0.05$. As expected, tasks are completed in smaller time fractions as the error tolerance grows. Although only considered as a factor here, future work may consider an application of Fitts's law for human movement (Fitts 1954) to characterize the role of error tolerance.

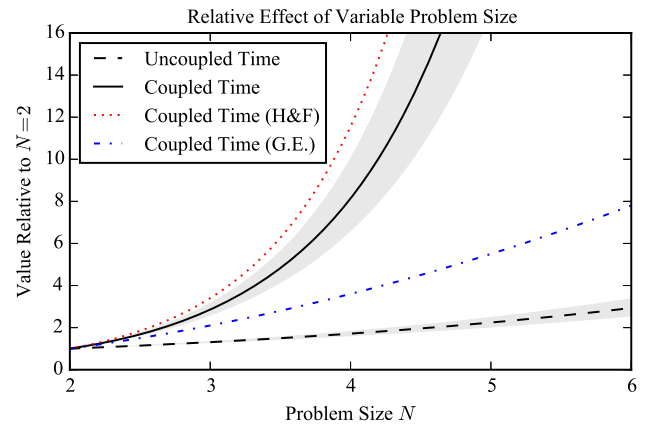


Fig. 6 Relative effect of problem size N on task completion time compared with results from H&F and Gaussian elimination (G.E.). Shaded area approximates a 95 % confidence interval on coefficient values

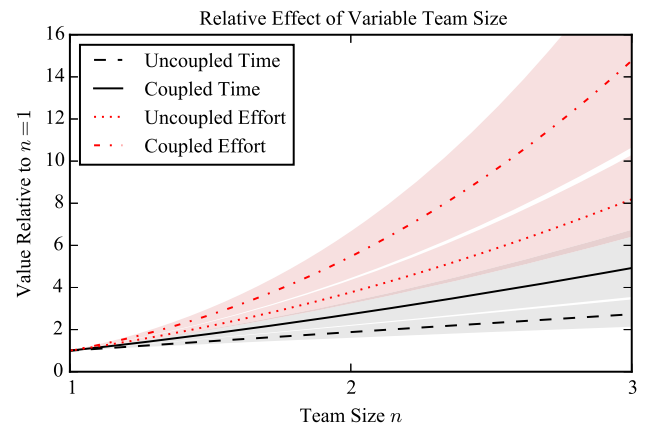


Fig. 7 Relative effect of team size n on task completion time and total effort (person-time). Shaded area approximates a 95 % confidence interval on coefficient values

Finally, the interaction term between problem and team size was found to be small and not significant for both uncoupled and coupled tasks. This surprising result may be due to the fixed organizational structures in this study, whereas real teams are structured to meet needs of the problem (e.g., Eppinger et al. 1994). Due to their apparent independence in this study, effects of technical and social complexity can be isolated in Figs. 6 and 7 relative to baseline cases. Figure 6 clearly illustrates the geometric and nearly linear growth for coupled and uncoupled problems as found in H&F. Figure 7 shows a smaller but distinct difference between coupled (super-linear) and uncoupled (sub-linear) tasks as team size grows. Note the costly 3–5 \times time multipliers for teams of three and corresponding 9–15 \times effort multiplier on total person-time. These results show high potential costs to collaboration—

independent of the technical problem—when designers are restricted by communication barriers.

6.2 Limitations

A few limitations of this study must be discussed in more detail. First, demographic factors such as age, gender, and frequency of previous interaction were largely omitted in analysis due to limited effects observed in preliminary study. While some of these factors likely influence outcomes as shown by Woolley et al. (2010), more data and refined instruments are required to distinguish their effects in aggregated groups. Future work with a broader sampling frame should carefully capture related factors of interest to identify differences in subject population.

This study's treatment of pairs (tasks 7–10) is imperfect in a three-subject session. Task 7 studies an uncoupled pair design task, while a third designer simultaneously completes an independent task. Tasks 8–10 provide a blank interface for the third designer. In both cases, the third subject in the room likely biases results to over-estimate effort due to additional conversation or distraction. Similarly, due to the structural similarities of pair and triad tasks, the group factor G aggregates across both rather than defining separate factors for each pair of designers. Future work may consider more controlled settings to study variable team sizes.

The restriction in six sessions preventing the $N_C = 4$ design tasks within the first ten tasks likely biases the results due to observed learning effects. Although controlled in analysis, limited samples of large tasks early in the experiment may under-estimate ordering effects. The one observed subject concession directly demonstrates this bias. Future work may consider more numerous training tasks to further decrease the effect of task ordering; however, practical considerations limit the fraction of time devoted to training versus data collection (about 1:4 here).

The decision to generate unique design tasks for each session introduces variance in results. Random generation of \mathbf{M} and \mathbf{y}^* may make certain tasks easier to solve in particular sessions, generally if some coupling factors m_{ij} are close to zero. While using the same set of tasks across all sessions would reduce this error, it also would introduce a wider bias to the specific set of tasks selected and limit repeatability of results. Future work may consider additional task replications to mitigate this effect.

Generalizability of results beyond the experimental frame is limited by the design scenario. As described in Sect. 3.5, the surrogate task models a context-free linear system with one zero-error solution where real design tasks are likely nonlinear, much larger (more variables), context-rich, and may have no or many solutions. Only small

ranges of problem size (2–4 variables) and team size (1–3 designers) are considered due to practical limitations. The surrogate task does not consider specialized knowledge (Postrel 2002) or decomposable tasks (Eppinger et al. 1994) which remove nearly all benefits from collaboration. The design team organizational structure is essentially fixed to be fully connected, while real product development networks exhibit the small-world property with local clustering but a short distance between any two individuals (Braha and Bar-Yam 2007). Furthermore, ad hoc team formation does not capture other factors relevant to group performance such as cohesion. Despite these limitations, results may be considered as stylized cases to bound features of real tasks.

Finally, subject selection from graduate engineering programs captures some aspects of the intended population (i.e., designers); however, students have varying degrees of experience and do not all practice design. This effect is mitigated by the context-free design task and ad hoc team composition which limit the impact of previous experience. Furthermore, the two-level regression model accommodates individual variance in computer interface manipulation and problem-solving capability by individuals and groups. Possible selection bias must be acknowledged from volunteer responses to email and convenience solicitation which demonstrates a general interest and degree of engagement in the design activity by the participants.

6.3 Implications for design

Although limited in generalizability, results from this study can bound some features of broader design tasks. Most real problems are partially coupled with a sparse \mathbf{M} matrix (Eppinger et al. 1994). Time to solve linear design variants is bounded by coupled and uncoupled problems in Fig. 6 which, together, form a lower bound on nonlinear design variants. The effect of context reduces effective technical complexity due to prior knowledge, suggesting these results are an upper bound on time to complete context-rich tasks.

Other efforts to reduce effective technical complexity may mitigate the geometric contribution of problem size in coupled problems. For example, higher information update frequency in the user interface may explain why results of this study improve upon H&F shown as a dotted line in Fig. 6. Gaussian elimination (G.E., shown as a dash-dot line) as an alternative solution method provides even better scaling; however, it requires centralized control and numerical methods purposefully difficult to implement in this scenario.

The effect of team size on completion time is similar for uncoupled and coupled tasks under the barriers to collaboration considered. Any value above 1.0 in Fig. 7 indicates

costs to collaboration as expected for non-decomposable tasks. Indeed, collaboration contributes about 50–60 % of time for pairs and 60–80 % for triads. These values are an upper bound to real design tasks which benefit from parallel work flows, rich communication, and specialized knowledge. One comparison point observes engineers spend less than half their time on “legitimate design acts” with the remainder devoted to resolving ambiguity, determining what must be known, and how to work with others (Bucciarelli 1988). The cost of collaboration is even higher when considering the total team effort incurred as person-time. Collaboration contributes 70–80 % of cost for pairs and 90 % for triads. Although these costs to collaboration cannot easily be generalized outside this study, an ideal collaborative process would maintain or improve upon the efficiency of an individual.

7 Conclusion

This study produces three main results. First, it replicates results from H&F with small but statistically significant differences in scaling factor for coupled tasks which may be attributed to contextual factors. Second, technical and social sources of complexity appear to be independent factors contributing to completion time for the tasks considered which allows the two factors to be studied in isolation. Finally, social sources of complexity from team size contribute significant time and cost in design under purposeful barriers to collaboration. Completion time grows sub-linearly with team size for uncoupled tasks and super-linearly for coupled tasks. These results and future efforts may improve estimates of required effort for collaborative design and help assess the effectiveness of proposed processes and tools.

There are several extensions of this study left for future work. First, external validity could be improved with revised design tasks. While the surrogate task in this study uses a linear system of equations, any general system model could be substituted. Nonlinear, context-rich, larger, or partially coupled system models would improve the generalizability of results. For example, De Jong functions (De Jong 1975) used in optimization evaluation may maintain a context-free task while introducing nonlinear models. Some system models may warrant an objective function to avoid a difficult-to-find single zero-error solution, a time limit for practical reasons, and both local and global objectives for realism.

Future studies may also benefit from larger, experienced, more familiar, organizationally diverse, or even spatially distributed design teams. Introducing partially coupled problems where the **M** matrix is neither complete nor diagonal allows tasks to be decomposed into sub-

problems. Alternative organizational structures may mirror the design problem (Eppinger et al. 1994), form a clustered network (Huberman and Hogg 1995), or reduce the level of highly connected designers (Braha and Bar-Yam 2007) to improve group performance. Additional analysis of actual communication (e.g., verbal transcriptions) may yield more details on information exchange between designers to compare optimal policies (Yassine et al. 2013) and evaluate assortivity or disassortivity (Braha et al. 2013).

Next, there may be opportunities to combine the uncoupled and coupled task evaluation under a common framework with a complexity metric. While treated as two separate classes of problems in this study, there are likely common underlying features which may be quantified using technical and social complexity metrics rather than the problem size N and team size n variables used in this analysis. One candidate is the comprehensive metric for structural complexity using graph energy as a measure of Shannon information entropy (Sinha 2014). Combined analysis would improve results by aggregating group factors across both coupled and uncoupled tasks and possibly allow extension to more generalizable problems.

Finally, another area of future work could assess proposed collaborative methods by comparing to the baseline case in this study. For example, sharing output values on a common display may allow uncoupled tasks to be completed in parallel. Other numerical outputs may help designers quantitatively evaluate the effect of input changes in complex design problems. Displaying the quantitative error between output and target values, normalized to a reasonable scale, could support collaborative decision-making by building a common mental model of “goodness.” This aligns with the “discourse group preference” phase of the engineering collaboration via negotiation (ECN) model hypothesized by Lu et al. (2007), which in the present study is limited to qualitative judgments.

Acknowledgments This research was completed with support from the US Department of Defense under a National Defense Science and Engineering Graduate (NDSEG) fellowship and with equipment support from the Center for Complex Engineering Systems at King Abdulaziz City for Science and Technology (KACST) and Massachusetts Institute of Technology. The authors thank Dan Frey for inspiring this study and helpful feedback along the way, Roi Guinto for his assistance in conducting experimental sessions, four anonymous reviewers for substantial feedback to improve this paper in the context of the design literature, and more than 30 participants for volunteering their time.

References

- Ameri F, Summers JD, Mocko GM, Porter M (2008) Engineering design complexity: an investigation of methods and measures. *Res Eng Des* 19(2–3):161–179. doi:[10.1007/s00163-008-0053-2](https://doi.org/10.1007/s00163-008-0053-2)

- Arena MV, Younossi O, Brancato K, Blockstein I, Grammich CA (2008) Why has the cost of fixed-wing aircraft risen? A macroscopic examination of the trends in U.S. military aircraft costs over the past several decades. RAND Corporation, Santa Monica
- Argote L, Insko CA, Yovetich N, Romero AA (1995) Group learning curves: the effects of turnover and task complexity on group performance. *J Appl Soc Psychol* 25(6):512–529. doi:[10.1111/j.1559-1816.1995.tb01765.x](https://doi.org/10.1111/j.1559-1816.1995.tb01765.x)
- Arias E, Eden H, Fisher G (1997) Enhancing communication, facilitating shared understanding, and creating better artifacts by integrating physical and computational media for design. In: Proceedings of the 2nd conference on designing interactive systems: processes, practices, methods, and techniques, pp 1–12. doi:[10.1145/263552.263553](https://doi.org/10.1145/263552.263553)
- Arias E, Eden H, Fischer G, Gorman A, Scharff E (2000) Transcending the individual human mind—creating shared understanding through collaborative design. *ACM T Comput Hum Int* 7(1):84–113. doi:[10.1145/344949.345015](https://doi.org/10.1145/344949.345015)
- Bashir HA, Thomson V (1999) Estimating design complexity. *J Eng Design* 10(3):247–257. doi:[10.1080/095448299261317](https://doi.org/10.1080/095448299261317)
- Bashir HA, Thomson V (2001) Models for estimating design effort and time. *Des Stud* 22(2):141–155. doi:[10.1016/S0142-694X\(00\)00014-4](https://doi.org/10.1016/S0142-694X(00)00014-4)
- Bates D, Mäechler M, Bolker BM, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48. doi:[10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
- Braha D, Bar-Yam Y (2007) The statistical mechanics of complex product development: empirical and analytical results. *Manage Sci* 53(7):1127–1145. doi:[10.1287/mnsc.1060.0617](https://doi.org/10.1287/mnsc.1060.0617)
- Braha D, Maimon O (1998a) A mathematical theory of design: foundations, algorithms and applications. Springer, Berlin. doi:[10.1007/978-1-4757-2872-9](https://doi.org/10.1007/978-1-4757-2872-9)
- Braha D, Maimon O (1998b) The measurement of a design structural and functional complexity. *IEEE T Syst Man Cyb* 28(4):527–535. doi:[10.1109/3468.686715](https://doi.org/10.1109/3468.686715)
- Braha D, Brown DC, Chakrabarti A, Dong A, Fadel G, Maier JR, Seering W, Ullman DG, Wood K (2013) DTM at 25: Essays on themes and future directions. In: ASME 2013 international design engineering technical conferences and computers and information in engineering conference, pp 1–17. doi:[10.1115/DETC2013-12280](https://doi.org/10.1115/DETC2013-12280)
- Briggs RO, De Vreede GJ, Nunamaker JF Jr (2003) Collaboration engineering with ThinkLets to pursue sustained success with group support systems. *J Manag Inf Syst* 19(4):31–64. doi:[10.1080/07421222.2003.11045743](https://doi.org/10.1080/07421222.2003.11045743)
- Bucciarelli LL (1988) An ethnographic perspective on engineering design. *Des Stud* 9(3):159–168. doi:[10.1016/0142-694X\(88\)90045-2](https://doi.org/10.1016/0142-694X(88)90045-2)
- Card D, Agresti W (1988) Measuring software design complexity. *J Syst Softw* 8(3):185–197. doi:[10.1016/0164-1212\(88\)90021-0](https://doi.org/10.1016/0164-1212(88)90021-0)
- Cohen SG, Bailey DE (1997) What makes teams work: group effectiveness research from the shop floor to the executive suite. *J Manag* 23(3):239–290. doi:[10.1177/014920639702300303](https://doi.org/10.1177/014920639702300303)
- Cooper KG, Lyneis JM, Bryant BJ (2002) Learning to learn, from past to future. *Int J Proj Manag* 20(3):213–219. doi:[10.1016/S0263-7863\(01\)00071-0](https://doi.org/10.1016/S0263-7863(01)00071-0)
- De Jong K (1975) An analysis of the behaviour of a class of genetic adaptive systems. Ph.D. thesis, University of Michigan, Ann Arbor
- Dirks KT (1999) The effects of interpersonal trust on work group performance. *J Apply Psychol* 84(3):445–455. doi:[10.1037/0021-9010.84.3.445](https://doi.org/10.1037/0021-9010.84.3.445)
- El-Haik B, Yang K (1999) The components of complexity in engineering design. *IIE Trans* 31(10):925–934. doi:[10.1080/07408179908969893](https://doi.org/10.1080/07408179908969893)
- Engelbart DC (1995) Toward augmenting the human intellect and boosting our collective IQ. *Commun ACM* 38(8):30–32. doi:[10.1145/208344.208352](https://doi.org/10.1145/208344.208352)
- Eppinger SD, Whitney DE, Smith RP, Gebala DA (1994) A model-based method for organizing tasks in product development. *Res Eng Des* 6(1):1–13. doi:[10.1007/BF01588087](https://doi.org/10.1007/BF01588087)
- Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol* 47(6):381–391. doi:[10.1037/h0055392](https://doi.org/10.1037/h0055392)
- Flager F, Gerver DJ, Kallman B (2014) Measuring the impact of scale and coupling on solution quality for building design problems. *Des Stud* 35(2):180–199. doi:[10.1016/j.destud.2013.11.001](https://doi.org/10.1016/j.destud.2013.11.001)
- Flyvbjerg B, Holm MS, Buhl S (2002) Underestimating costs in public works projects: error or lie? *J Am Plann Assoc* 68(3):279–295. doi:[10.1080/01944360208976273](https://doi.org/10.1080/01944360208976273)
- Friedenthal S, Moore A, Steiner R (2012) A practical guide to SysML, the systems modeling language, 2nd edn. Elsevier, Waltham
- Gentle JE (1998) Numerical linear algebra for applications in statistics. Springer, New York, pp 87–121
- Goodman PS, Leyden DP (1991) Familiarity and group productivity. *J Apply Psychol* 76(4):578–586. doi:[10.1037/0021-9010.76.4.578](https://doi.org/10.1037/0021-9010.76.4.578)
- Halekoh U, Højsgaard S (2014) A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbkrtest. *J Stat Softw* 59(9):1–32. doi:[10.18637/jss.v059.i09](https://doi.org/10.18637/jss.v059.i09)
- Haskins C (ed) (2011) Systems engineering handbook. INCOSE-TP-2003-002-03.2.2, San Diego
- Henderson B (1968) The experience curve. *Perspectives* 16:1–2
- Hirschi NW, Frey DD (2002) Cognition and complexity: an experiment on the effect of coupling in parameter design. *Res Eng Des* 13:123–131. doi:[10.1007/s00163-002-0011-3](https://doi.org/10.1007/s00163-002-0011-3)
- Huberman BA, Hogg T (1995) Communities of practice: performance and evolution. *Comput Math Org Theory* 1(1):73–92. doi:[10.1007/BF01307829](https://doi.org/10.1007/BF01307829)
- Ingham AG, Levinger G, Graves J, Peckham V (1974) The Ringelmann effect: studies of group size and group performance. *J Exp Soc Psychol* 10(4):371–384. doi:[10.1016/0022-1031\(74\)90033-X](https://doi.org/10.1016/0022-1031(74)90033-X)
- Jehn KA, Pradhan P (1997) Interpersonal relationships and task performance: an examination of mediation processes in friendship and acquaintance groups. *J Pers Soc Psychol* 72(4):775–790. doi:[10.1037/0022-3514.72.4.775](https://doi.org/10.1037/0022-3514.72.4.775)
- Kearns M, Suri S, Montfort N (2006) An experimental study of the coloring problem on human subject networks. *Science* 313:824–827. doi:[10.1126/science.1127207](https://doi.org/10.1126/science.1127207)
- Kerr NL, Tindale RS (2004) Group performance and decision making. *Annu Rev Psychol* 55:623–655. doi:[10.1146/annurev.psych.55.090902.142009](https://doi.org/10.1146/annurev.psych.55.090902.142009)
- Konda S, Monarch I, Sargent P, Subrahmanian E (1992) Shared memory in design: a unifying theme for research and practice. *Res Eng Des* 4(1):23–42. doi:[10.1007/BF02032390](https://doi.org/10.1007/BF02032390)
- Kravitz DA, Martin B (1986) Ringelmann rediscovered: the original article. *J Pers Soc Psychol* 50(5):936–941. doi:[10.1037/0022-3514.50.5.936](https://doi.org/10.1037/0022-3514.50.5.936)
- Lazer D, Friedman A (2007) The network structure of exploration and exploitation. *Admin Sci Q* 52(4):667–694. doi:[10.2189/asqu.52.4.667](https://doi.org/10.2189/asqu.52.4.667)
- Lu SCY, Elmaraghy W, Schuh G, Wilhelm R (2007) A scientific foundation of collaborative engineering. *CIRP Ann Manuf Technol* 56(2):605–634. doi:[10.1016/j.cirp.2007.10.010](https://doi.org/10.1016/j.cirp.2007.10.010)
- Lyneis JM, Cooper KG, Els SA (2001) Strategic management of complex projects: a case study using system dynamics. *Syst Dyn Rev* 17(3):237–260. doi:[10.1002/sdr.213](https://doi.org/10.1002/sdr.213)

- Maier MW (1998) Architecting principles for systems-of-systems. *Syst Eng* 1(4):267–284
- Mason W, Watts DJ (2012) Collaborative learning in networks. *P Natl Acad Sci USA* 109(3):764–769. doi:[10.1073/pnas.1110069108](https://doi.org/10.1073/pnas.1110069108)
- McCubbins MD, Paturi R, Weller N (2009) Connected coordination: network structure and group coordination. *Am Polit Res* 37(5):899–920. doi:[10.1177/1532673X09337184](https://doi.org/10.1177/1532673X09337184)
- Miller GA (1956) The magic number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 63:81–97. doi:[10.1037/h0043158](https://doi.org/10.1037/h0043158)
- Moløkken K, Jørgensen M (2003) A review of surveys on software effort estimation. In: *Proceedings of the 2003 ACM-IEEE international symposium on empirical software engineering*, pp 220–230. doi:[10.1109/ISESE.2003.1237981](https://doi.org/10.1109/ISESE.2003.1237981)
- Mullen B, Copper C (1994) The relation between group cohesiveness and performance: an integration. *Psychol Bull* 115(2):210–227. doi:[10.1037/0033-2909.115.2.210](https://doi.org/10.1037/0033-2909.115.2.210)
- National Aeronautical and Space Administration (2007) *NASA systems engineering handbook*. NASA/SP-2007-6105 Rev1, Washington, D.C
- National Research Council (2010) *Controlling cost growth of NASA earth and space science missions*. Washington, D.C
- Postrel S (2002) Islands of shared knowledge: specialization and mutual understanding in problem-solving teams. *Org Sci* 13(3):303–320. doi:[10.1287/orsc.13.3.303.2773](https://doi.org/10.1287/orsc.13.3.303.2773)
- Raudenbush SW, Bryk AS (2002) *Hierarchical linear models: applications and data analysis methods*, 2nd edn. Sage Publications, Thousand Oaks
- Rhodes DH, Ross AM (2010) Five aspects of engineering complex systems: Emerging constructs and methods. In: *4th Annual IEEE systems conference*, San Diego. doi:[10.1109/SYSTEMS.2010.5482431](https://doi.org/10.1109/SYSTEMS.2010.5482431)
- Sage AP, Cuppan CD (2001) On the systems engineering and management of systems of systems and federations of systems. *Information knowledge systems management* 2(4):325–345
- Shore J, Bernstein E, Lazer D (2015) Facts and figuring: an experimental investigation of network structure and performance in information and solution spaces. *Org Sci* 26(5):1432–1446. doi:[10.1287/orsc.2015.0980](https://doi.org/10.1287/orsc.2015.0980)
- Sinha K (2014) *Structural complexity and its implications for design of cyber-physical systems*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge
- Smith RP, Eppinger SD (1997) Identifying controlling features of engineering design iteration. *Manag Sci* 43(3):276–293. doi:[10.1287/mnsc.43.3.276](https://doi.org/10.1287/mnsc.43.3.276)
- Suh NP (1999) A theory of complexity, periodicity and the design axioms. *Res Eng Des* 11(2):116–132. doi:[10.1007/PL00003883](https://doi.org/10.1007/PL00003883)
- Summers JD, Shah JJ (2010) Mechanical engineering design complexity metrics: size, coupling, and solvability. *J Mech Design* 132(2):021004. doi:[10.1115/1.4000759](https://doi.org/10.1115/1.4000759)
- US Government Accountability Office (2011) *Trends in Nunn-McCurdy breaches for major defense acquisition programs*. GAO-11-295R, Washington, D.C
- Wang L, Shen W, Xie H, Neelamkavil J, Pardasani A (2002) Collaborative conceptual design-state of the art and future trends. *Comput Aided Design* 34(13):981–996. doi:[10.1016/S0010-4485\(01\)00157-9](https://doi.org/10.1016/S0010-4485(01)00157-9)
- Weingart LR (1992) Impact of group goals, task component complexity, effort, and planning on group performance. *J Apply Psychol* 77(5):682–693. doi:[10.1037/0021-9010.77.5.682](https://doi.org/10.1037/0021-9010.77.5.682)
- Williams T (2005) Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE T Eng Manage* 52(4):497–508. doi:[10.1109/TEM.2005.856572](https://doi.org/10.1109/TEM.2005.856572)
- Wood RE (1986) Task complexity: definition of the construct. *Org Behav Hum Decis* 37:60–82. doi:[10.1016/0749-5978\(86\)90044-0](https://doi.org/10.1016/0749-5978(86)90044-0)
- Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330:686–688. doi:[10.1126/science.1193147](https://doi.org/10.1126/science.1193147)
- Yassine A, Joglekar N, Braha D, Eppinger S, Whitney D (2003) Information hiding in product development: the design churn effect. *Res Eng Des* 14(3):145–161. doi:[10.1007/s00163-003-0036-2](https://doi.org/10.1007/s00163-003-0036-2)
- Yassine A, Maddah B, Nehme N (2013) Optimal information exchange policies in integrated product development. *IIE Trans* 45(12):1249–1262. doi:[10.1080/0740817X.2012.762487](https://doi.org/10.1080/0740817X.2012.762487)