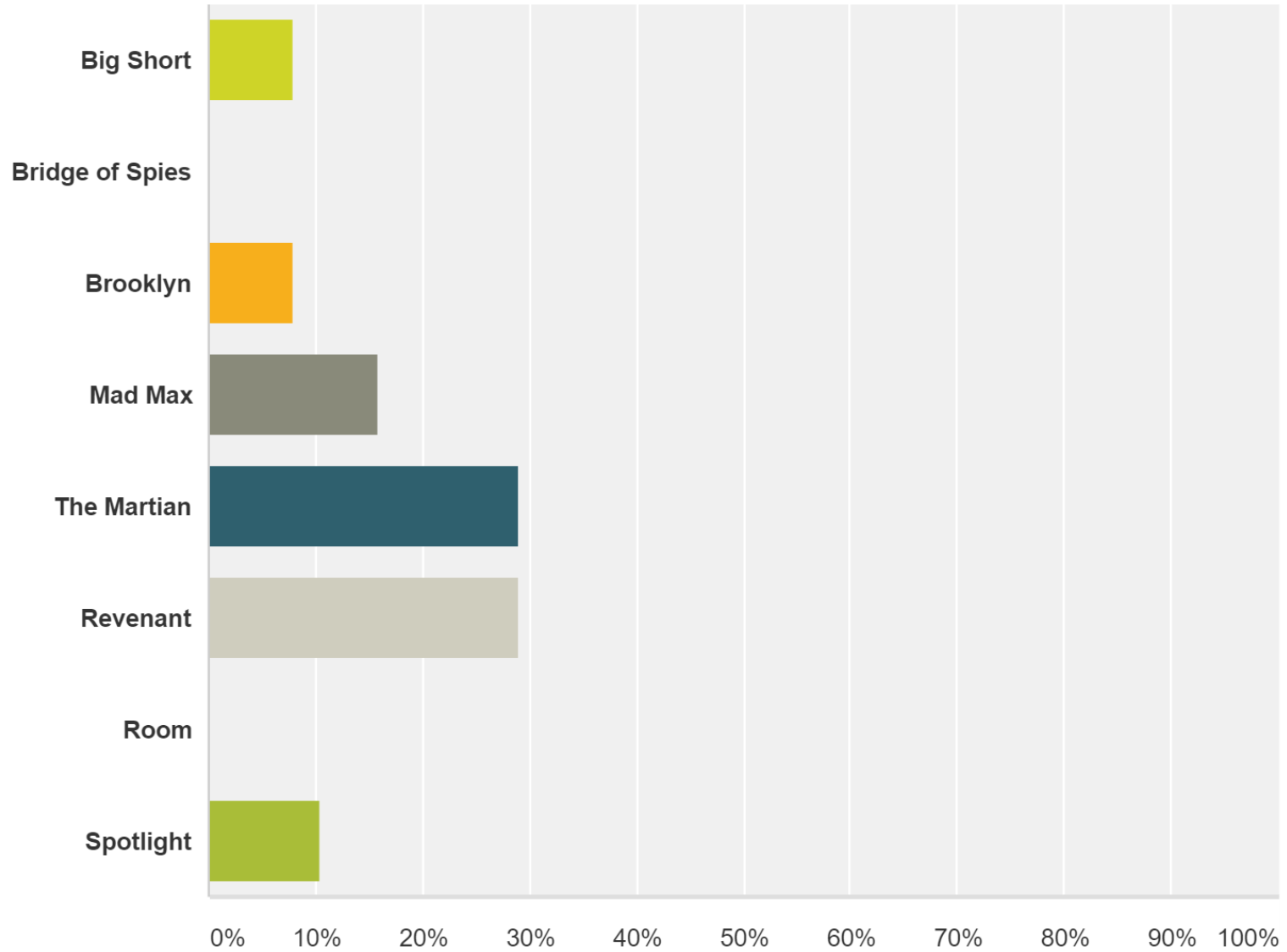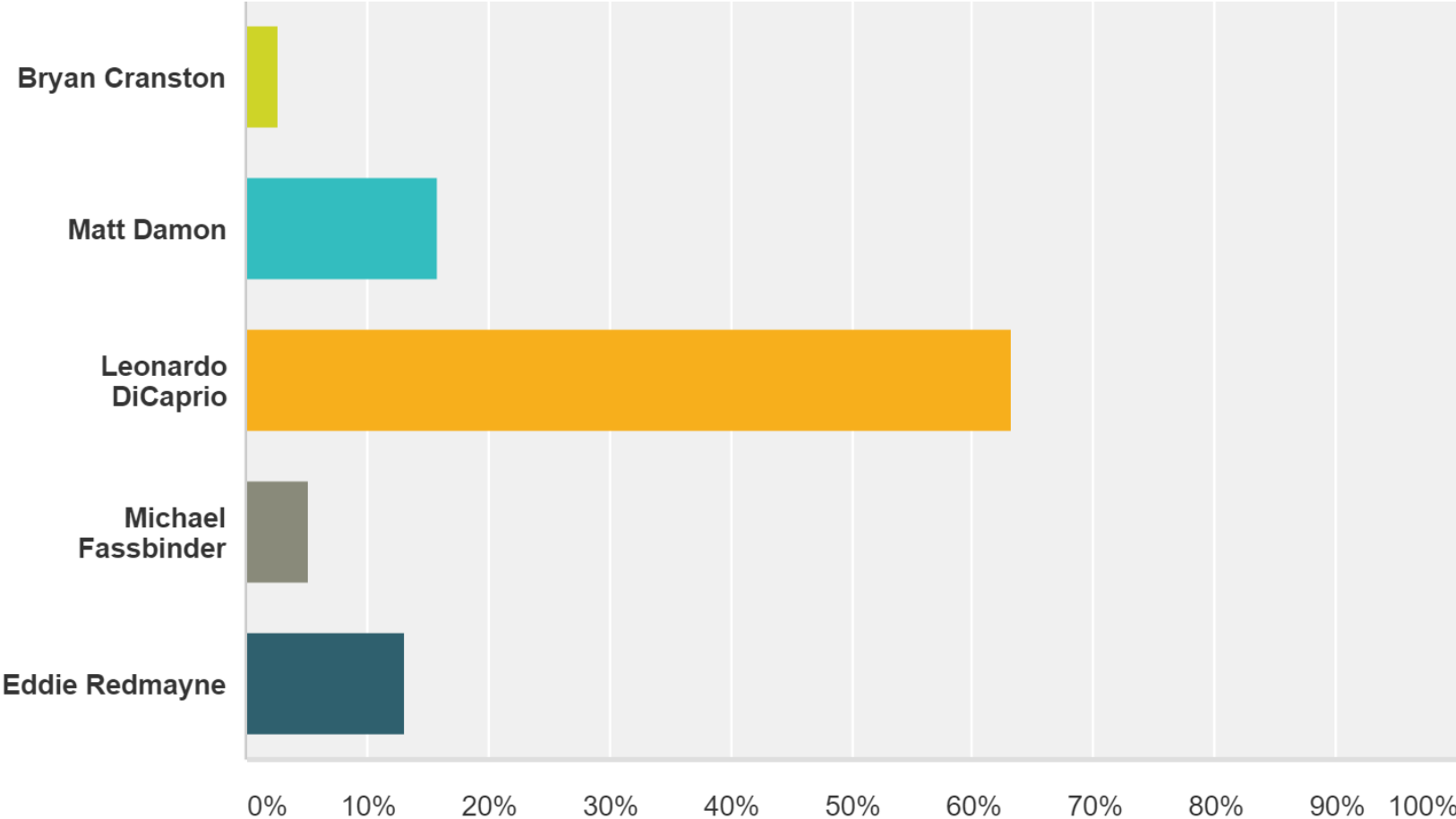# Lecture 9:
CS 5306 / INFO 5306:
Crowdsourcing and
Human Computation

Which movie will win for Best Picture?

Answered: 38   Skipped: 0
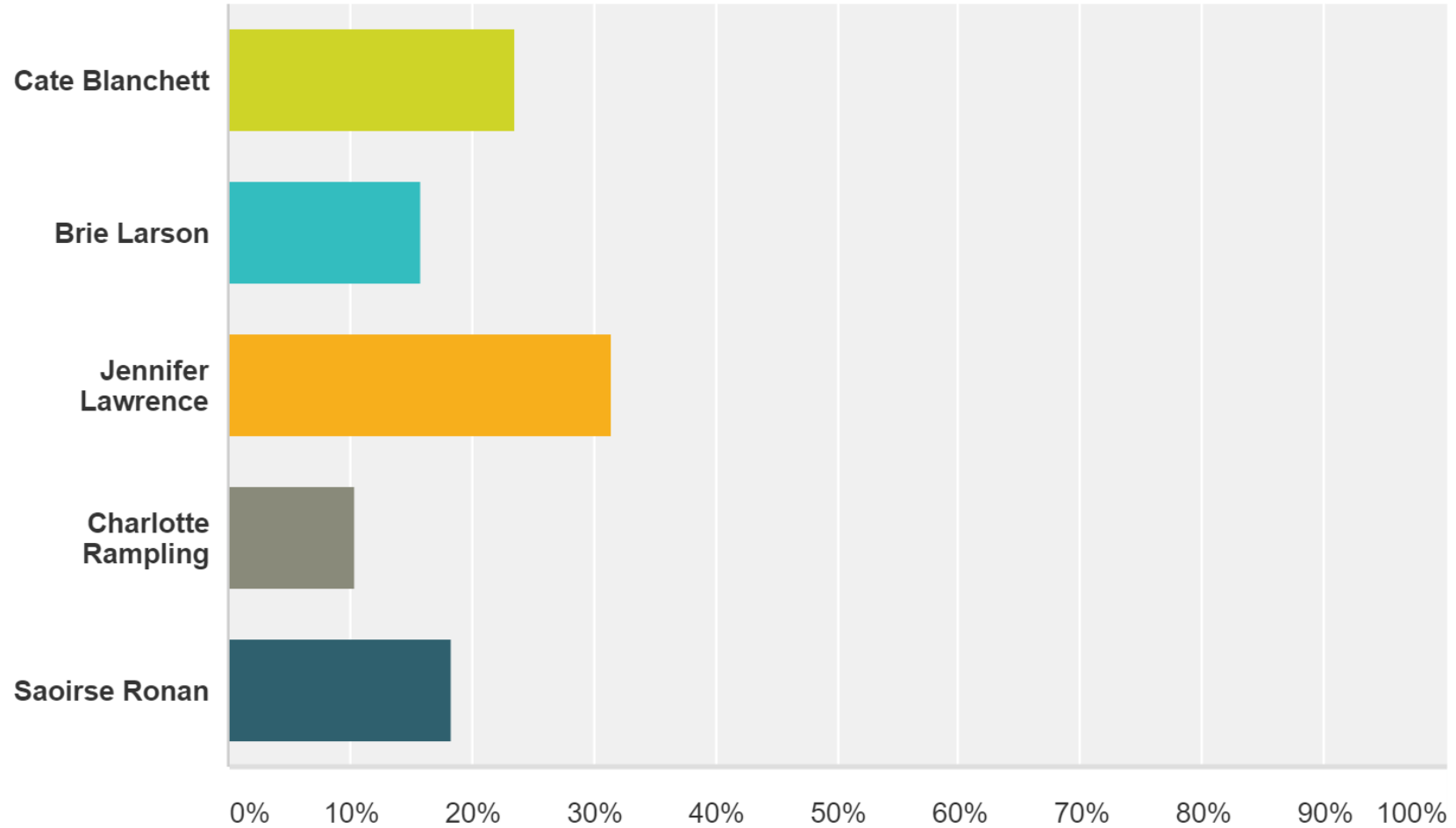
Who will win Best Actor

Who will win Best Actress?

Answered: 38   Skipped: 0

Amazon Mechanical Turk Accounts

1. Fill out web page to request becoming an Mturk worker
   - Requires either a social security number (SSN) or an individual tax identification number (ITIN).

2. Contact MTurk support at contactus@mturk.com using the email address that you used to sign up for the MTurk account
   - Tell them the account is for your Cornell Crowdsourcing and Human Computation course
   - Following instructions given by Amy J at Mturk to course TA Moontae Lee

- Please do this within the next 48 hours
- If you have problems let me know

Aristotle, *Politics*, Book 3 1281

The other alternatives may be reserved for a later inquiry; but the first of the alternatives suggested—that the people at large should be sovereign rather than the few best—would appear to be defensible, and while it presents some difficulty it perhaps also contains some truth. There is this to be said for the Many. Each of them by himself may not be of a good quality; but when they <u>all come together</u> it is possible that <u>they may surpass—collectively and as a body. although not individually—the quality of the few best</u>. Feasts to which many contribute may excel those provided at one man's expense. In the same way, <u>when there are many who contribute to the process of deliberation each can bring his share of goodness and moral prudence</u>; and when all meet together the people may thus become something in the nature of a single person who—as he has many feet, many hands, and many senses—may also have many qualities of character and intelligence. This is the reason why the Many are also better judges than the few of music and the writing of poets: <u>some appreciate one part, some another, and all together appreciate all</u>.

Aristotle, *Politics*, Book 3 1281

We may note that this combination of qualities, which gives the Many their merit, can also be traced in cases of *individual* merit. The thing which makes a good man differ from a unit in the crowd—as it is also the thing which is generally said to make a beautiful person differ from one who is not beautiful, or an artistic representation differ from ordinary reality—is that elements which are elsewhere scattered and separate are here combined in a unity. It is this unity which counts; for if you take the elements separately, you may say of an artistic representation that it is surpassed by the eye of this person or by some other feature of that.

Aristotle, *Politics*, Book 3 1281

It is not clear, however, that this combination of qualities, which we have made the ground of distinction between the many and the few best, is true of all popular bodies and all large masses of men. Perhaps it may be said, "By heaven, it is clear that there are some bodies of which it cannot possibly be true; for if you included them, you would, by the same token, be found to include a herd of beasts. That would be absurd; and yet what difference is there between these bodies and a herd of beasts?" All the same, and in spite of this objection, there is nothing to prevent the view we have stated from being true of *some* popular bodies.

John Rawls, *A Theory of Justice*

Nevertheless, we normally assume that an ideally conducted discussion among many persons is more likely to arrive at the correct conclusion (by a vote if necessary) than the deliberations of any one of the them by himself. Why should this be so? In everyday life the exchange of opinion with others checks our partiality and widens our perspective; we are made to see things from their standpoint and the limits of our vision are brought home to us. But in the ideal process the veil of ignorance means that the legislators are already impartial. <u>The benefits from the discussion lie in the fact that even representative legislators are limited in knowledge and ability to reason. No one of them knows everything the others know, or can make all the same inferences that they can draw in concert. Discussion is a way of combining information and englargeing the range of arguments.</u> At least in the course of time, the effects of common deliberation seem bound to improve matters.

Deliberation Seems Like a Good Thing

- Get information from the smartest person

- Aggregate distributed information

- Synergy and learning

Deliberation Often Isn't a Good Thing

- Deliberating groups are no better than statistical groups

Deliberation Often Isn't a Good Thing

- Deliberation gives false sense of security about decisions

Deliberation Often Isn't a Good Thing

- Groups members feel "majority pressure"
  - Informational influences:
    if everyone else agrees, perhaps I'm wrong
  - Social influences:
    I want to be liked

  Can be framed economically – private benefits vs social benefits

Deliberation Often Isn't a Good Thing

- Group members of "low social status" (in appropriate circumstances less educated people , women) speak less and have less influence in the group compared to higher-status members

task (correct, error, inserted error, and corrected error) to allow typists to distinguish sources of errors and correct responses and, therefore, provide a stronger test of illusions of authorship. We asked 24 skilled typists (WPM = 70.7 ± 16.4) to type 600 words, each of which was followed by a four-alternative explicit report screen. Typists typed 91.8% of the words correctly. Mean interkeystroke intervals, plotted in Fig. 3A, show post-error slowing for incorrect responses ($F_{1,18} = 117.7$, $p < 0.01$) and corrected errors ($F_{1,18} = 120.0$, $p < 0.01$), but not for inserted errors ($F < 1.0$), indicating that inner-loop detection distinguishes between actual errors and correct responses.

Explicit detection probabilities, plotted in Fig. 3B, show good discrimination between correct and error responses. For correct responses, typists said "correct" more than "error" [$a(23) = 97.29$, $p < 0.01$]; for error responses, typists said "error" more than "correct" [$a(23) = 8.22$, $p < 0.01$]. Typists distinguished actual errors from inserted errors well, avoiding an illusion of authorship. They said "error" more than "inserted" for actual errors [$a(23) = 7.06$, $p < 0.01$] and "inserted" more than "error" for inserted errors [$a(23) = 14.75$, $p < 0.01$]. However, typists showed a strong illusion of authorship with corrected errors. They were just as likely to call them correct responses as corrected errors [$a(23) = 1.38$].

The post-error slowing and post-trial report data show a dissociation between inner- and outer-loop error detection. We assessed the dissociation further by comparing post-error slowing on trials in which typists did and did not experience illusions of authorship (21). The pattern of post-error slowing was the same for both sets of trials (Fig. S6), suggesting that the pattern in Fig. 3A is representative of all trials.

The three experiments found strong dissociations between explicit error reports and post-error slowing. These dissociations are consistent with the hierarchical error-detection mechanism that we proposed, with an outer loop that mediates explicit reports and an inner loop that mediates post-error slowing. This nested-loop description of error detection is consistent with hierarchical models of cognitive control in typewriting (9, 10, 15–17) and with models of hierarchical control in other complex tasks (2, 8, 22). Speaking, playing music, and navigating through space may all involve inner loops that take care of the details of performance (e.g., uttering phonemes, playing notes, and walking) and outer loops that ensure that intentions are fulfilled (e.g., messages communicated, songs performed, and destinations reached). Hierarchical control may be prevalent in highly skilled performers who have had enough practice to develop an autonomous inner loop. Previous studies of error detection in simple tasks may describe inner-loop processing. The novel contribution of our research is to dissociate the outer loop from the inner loop.

The three experiments demonstrate cognitive illusions of authorship in skilled typewriting (11–14). Typists readily take credit for correct output on the screen, interpreting corrected errors as their own correct responses. They take the blame for inserted errors, as in the first and second experiments, but they also blame the computer, as in the third experiment. These illusions are consistent with the hierarchical model of error detection, with the outer loop assigning credit and blame and the inner loop doing the work of typing (10, 17). Thus, illusions of authorship may be a hallmark of hierarchical control systems (2, 11, 22, 23).

References and Notes
1. P. M. A. Rabbitt, J. Exp. Psychol. 71, 264 (1966).
2. D. A. Norman, Psychol. Rev. 88, 1 (1981).
3. C. B. Holroyd, M. G. H. Coles, Psychol. Rev. 109, 679 (2002).
4. N. Yeung, M. M. Botvinick, J. D. Cohen, Psychol. Rev. 111, 931 (2004).
5. W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, E. Donchin, Psychol. Sci. 4, 385 (1993).
6. S. Dehaene, M. I. Posner, D. M. Tucker, Psychol. Sci. 5, 303 (1994).
7. C. S. Carter et al., Science 280, 747 (1998).
8. K. S. Lashley, in Cerebral Mechanisms in Behavior, L. A. Jeffress, Ed. (Wiley, New York, 1951), pp. 112–136.
9. T. A. Salthouse, Psychol. Bull. 99, 303 (1986).
10. G. D. Logan, M. J. C. Crump, Psychol. Sci. 20, 1296 (2009).
11. T. I. Nielsen, Scand. J. Psychol. 4, 225 (1963).
12. M. M. Botvinick, J. D. Cohen, Nature 391, 756 (1998).
13. D. M. Wegner, The Illusion of Conscious Will (MIT Press, Cambridge, MA, 2002).
14. G. Knoblich, T. T. J. Kircher, J. Exp. Psychol. Hum. Percept. Perform. 30, 657 (2004).
15. D. E. Rumelhart, D. A. Norman, Cogn. Sci. 6, 1 (1982).
16. L. H. Shaffer, Psychol. Rev. 83, 375 (1976).
17. X. Liu, M. J. C. Crump, G. D. Logan, Mem. Cognit. 38, 474 (2010).
18. A. M. Gordon, J. F. Soechting, Exp. Brain Res. 107, 281 (1995).
19. J. Long, Ergonomics 19, 93 (1976).
20. P. Rabbitt, Ergonomics 21, 945 (1978).
21. Materials and methods are available as supporting material on Science Online.
22. M. M. Botvinick, Trends Cogn. Sci. 12, 201 (2008).
23. R. Cooper, T. Shallice, Cogn. Neuropsychol. 17, 297 (2000).
24. We thank J. D. Schall for comments on the manuscript. This research was supported by grants BCS 0646588 and BCS 0957074 from the NSF.

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

Anita Williams Woolley,[1]* Christopher F. Chabris,[2,4] Alex Pentland,[3,4] Nada Hashmi,[4,5] Thomas W. Malone[4,5]

Psychologists have repeatedly shown that a single statistical factor—often called "general intelligence"—emerges from the correlations among people's performance on a wide variety of cognitive tasks. But no one has systematically examined whether a similar kind of "collective intelligence" exists for groups of people. In two studies with 699 people, working in groups of two to five, we find converging evidence of a general collective intelligence factor that explains a group's performance on a wide variety of tasks. This "c factor" is not strongly correlated with the average or maximum individual intelligence of group members but is correlated with the average social sensitivity of group members, the equality in distribution of conversational turn-taking, and the proportion of females in the group.

As research, management, and many other kinds of tasks are increasingly accomplished by groups—working both face-to-face and virtually (1–3)—it is becoming ever more important to understand the determinants of group performance. Over the past century, psychologists made considerable progress in defining and systematically measuring intelligence in individuals (4). We have used the statistical approach they developed for individual intelligence to systematically measure the intelligence of groups. Even though social psychologists and others have studied for decades how well groups perform specific tasks (5, 6), they have not attempted to measure group intelligence in the same way individual intelligence is measured—by assessing how well a single group can perform a wide range of different tasks and using that information to predict how that same group will perform other tasks in the future. The goal of the research reported here was to test the hypothesis that groups, like individuals, do have characteristic levels of intelligence, which can be measured and used to predict the groups' performance on a wide variety of tasks.

Although controversy has surrounded it, the concept of measurable human intelligence is based on a fact that is still as remarkable as it was to Spearman when he first documented it in 1904

[1]Carnegie Mellon University, Tepper School of Business, Pittsburgh, PA 15213, USA. [2]Union College, Schenectady, NY 12308, USA. [3]Massachusetts Institute of Technology (MIT) Center for Collective Intelligence, Cambridge, MA 02142, USA. [4]MIT Media Lab, Cambridge, MA 02139, USA. [5]MIT Sloan School of Management, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: awoolley@cmu.edu

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

- Intelligence: "People who do well on one mental task tend to do well on most others, despite large variations in the tests' contents and methods of administration."

Evidence for a Collective Intelligence Factor in the Performance of Human Groups
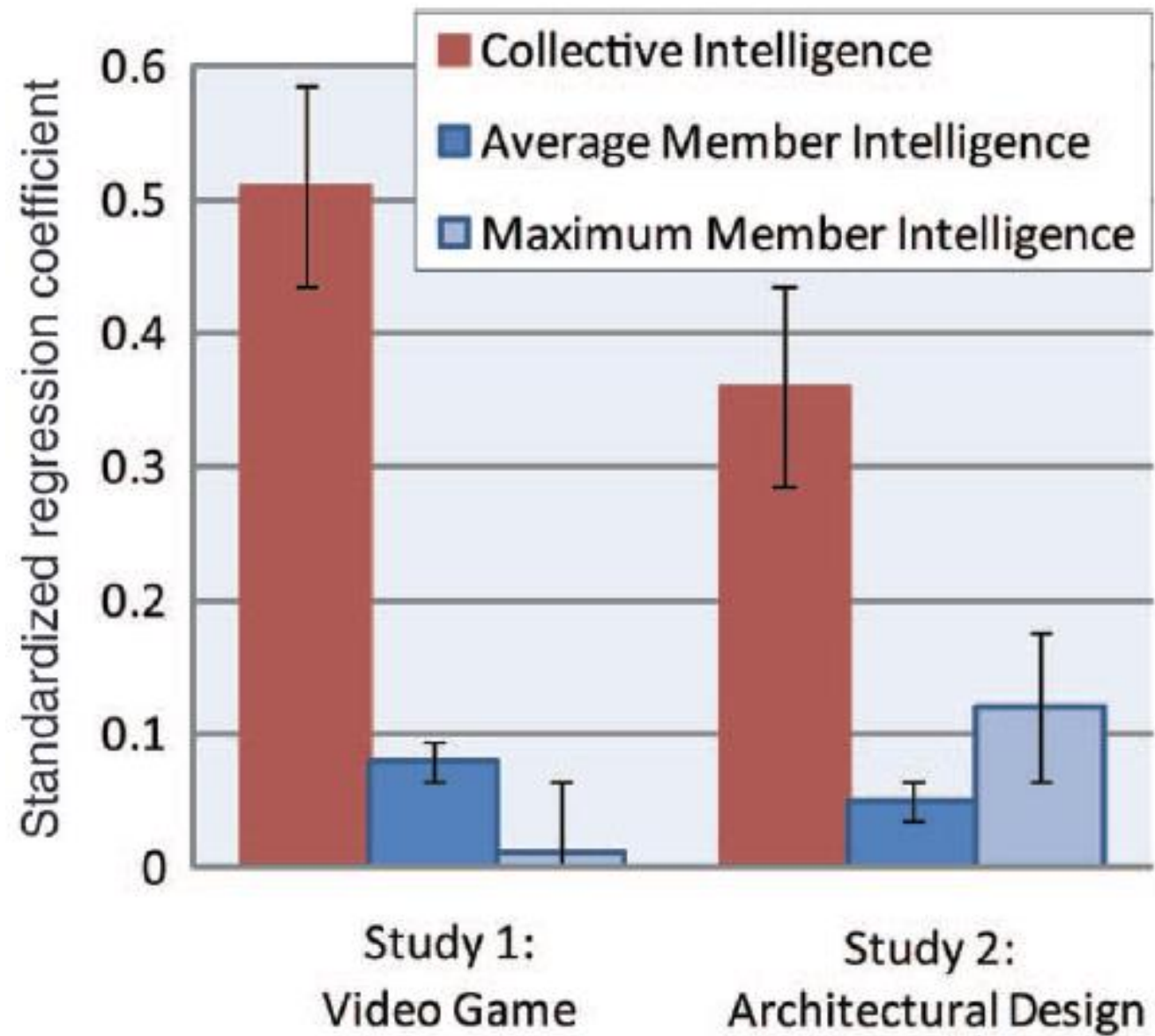
- Collective Intelligence factor (c):
  - Analogous to human IQ – "the general ability of the group to perform a wide variety of tasks
  - A group that performs well on a set of tasks are more likely to perform well on other tasks

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

- Their results show that there is a single dominant factor c in group performance
- Not strongly connected to average or maximum member intelligence, nor to group cohesion, motivation, or satisfaction

Other Measures

- *Group satisfaction.* Agreement with statements such as "I have been very satisfied working on this team".

- *Motivation.* Agreement with statements such as "I would feel bad and unhappy if our team has performed poorly".

- *Social cohesiveness.* Agreement with statements such as "Members of this group would enjoy being at a party together".

- *Psychological Safety.* Agreement with statements such as "It is difficult to ask other members of this team for help".

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

- It *was* significantly correlated with:
  - Average social sensitivity of group members
    - Used "Reading the Mind in the Eyes" test "
  - More turn taking
  - Proportion of females
    - Known to exhibit better social sensitivity and turn taking

Reading the Mind in the Eyes

Sociometric Badges

McGrath Task Circumplex

Quadrant I
Generate

Generating ideas

Generating plans

Solving problems with correct answers

Type 2: Creativity tasks

Type 1: Planning tasks

Executing performance tasks

Type 3: Intellective tasks

Type 8: Performances/ psycho-motor tasks

Cooperation

Quadrant II
Choose

Quadrant IV
Execute

Type 4: Decision-making tasks

Type 7: Contests/battles/ competitive tasks

Deciding issues with no right answer

Resolving conflicts of power

Conflict

Type 5: Cognitive conflict tasks

Type 6: Mixed-motive tasks

Resolving conflicts of viewpoint

Resolving conflicts of interest

Quadrant III
Negotiate

Conceptual

Behavioral

- *Quadrant I.*
  - *Brainstorming.* 10 minutes brainstorming possible uses for a brick.
  - *Word completions.* 10 minutes to come up with words beginning with "s" and ending with "n".
- *Quadrant II.*
  - *Group Matrix Reasoning.* Raven's Advanced Progressive Matrices.
  - S*patial problems .* 10 minutes generating ways to fit 6 3D rectangles into a 3D container.
  - *Group Moral Reasoning.* Decide on disciplinary actions in a case where a college basketball player bribed an instructor to change his grade on an exam.
  - *Incomplete words.* 10 minutes to complete a set of 36 words with 2-3 letters missing (" _ ech_ _ que" / "technique").
  - *Estimation Problems.* 10 minutes to estimate 20 quantities ("What was the median age in the U.S. in 2009?")

- *Quadrant III.*
  - *Plan shopping trip*. Plan a shopping trip as if they were all residents of the same house sharing a single car. Each member had a different list of groceries and various constraints such as better and worse places to buy the different items.
- *Quadrant IV.*
  - *Group typing.* 10 minutes to simultaneously type into a shared online document.
  - *Reproducing art*. Duplicate a hard copy of a picture created by coloring cells in a spreadsheet, the picture as accurately as possible using a shared online spreadsheet tool

- *Criterion tasks:*
  - *Video checkers.* +1 point for each move, +2 for each piece captured, +3 for each king.
  - *Architectural design*. Design and build a house, garage, and pool out of a limited set of building blocks subject to constraints

Raven's Advanced Progressive Matrices
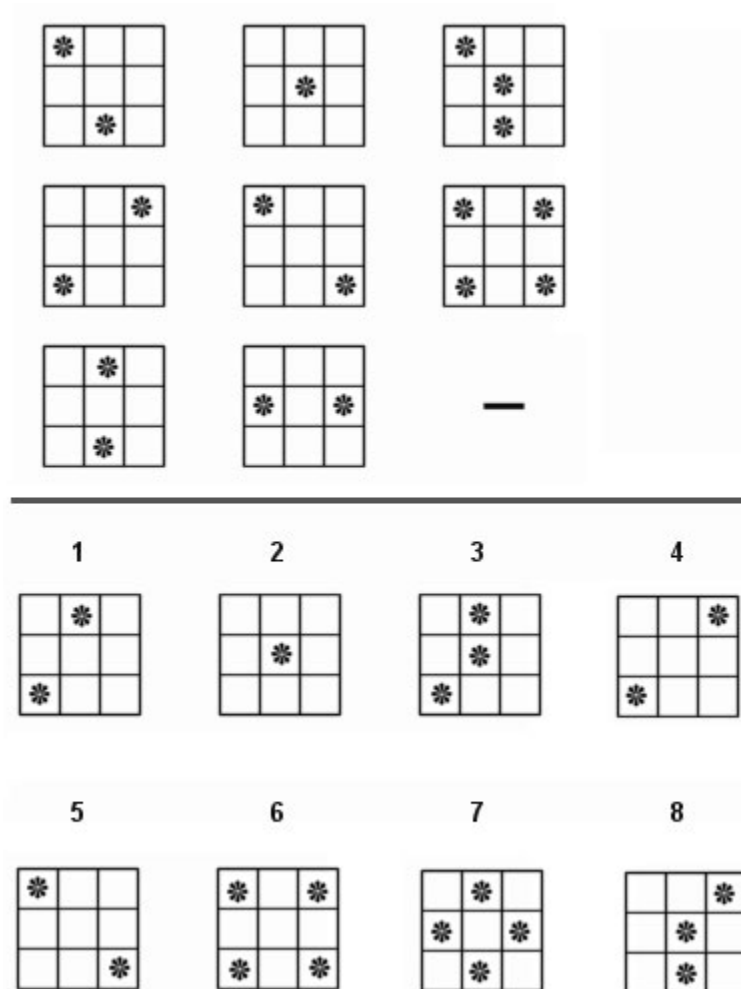
NEO Personality Inventory
(The "Big Five" Personality Traits)

- Neuroticism
  - Anxiety
  - Hostility
  - Depression
  - Self-Consciousness
  - Impulsiveness
  - Vulnerability to Stress
- Extraversion
  - Warmth
  - Gregariousness
  - Assertiveness
  - Activity
  - Excitement Seeking
  - Positive Emotion
- Openness to experience
  - Fantasy
  - Aesthetics
  - Feelings
  - Actions
  - Ideas
  - Values

- Agreeableness
  - Trust
  - Straightforwardness
  - Altruism
  - Compliance
  - Modesty
  - Tendermindedness
- Conscientiousness
  - Competence
  - Order
  - Dutifulness
  - Achievement Striving
  - Self-Discipline
  - Deliberation

Wonderlic Personnel Test

Which of the following is the earliest date?

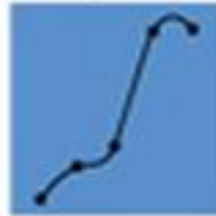   A) Jan. 16, 1898   B) Feb. 21, 1889   C) Feb. 2, 1898   D) Jan. 7, 1898   E) Jan. 30, 1889

LOW is to **HIGH** as **EASY** is to ___?___.

    J) **SUCCESSFUL**   K) **PURE**   L) **TALL**   M) **INTERESTING**   N) **DIFFICULT**

A featured product from an Internet retailer generated 27, 99, 80, 115 and 213 orders over a 5-hour period. Which graph below best represents this trend?
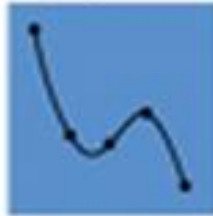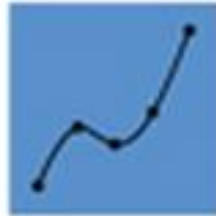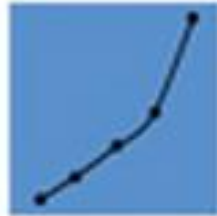

    A   B   C   D   E

What is the next number in the series?   29   41   53   65   77   ___?___

    J) 75   K) 88   L) 89   M) 98   N) 99

One word below appears in color. What is the OPPOSITE of that word?
She gave a complex answer to the question and we all agreed with her.

    A) long   B) better   C) simple   D) wrong   E) kind

"Four Big Problems" for Deliberating Groups

- Amplifying ("architectural") errors
- Hidden profiles and (favoring) common knowledge
- Cascades and polarization
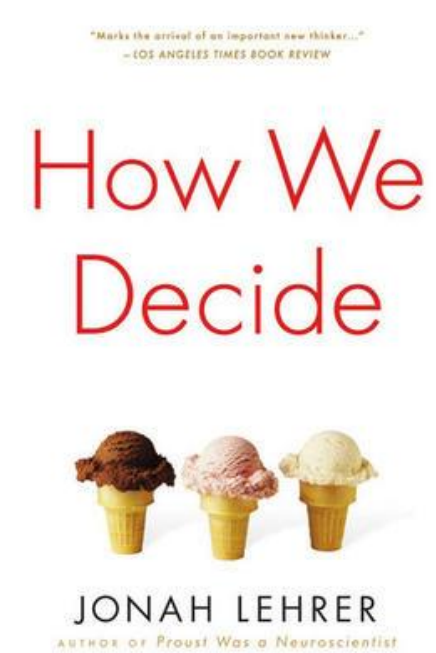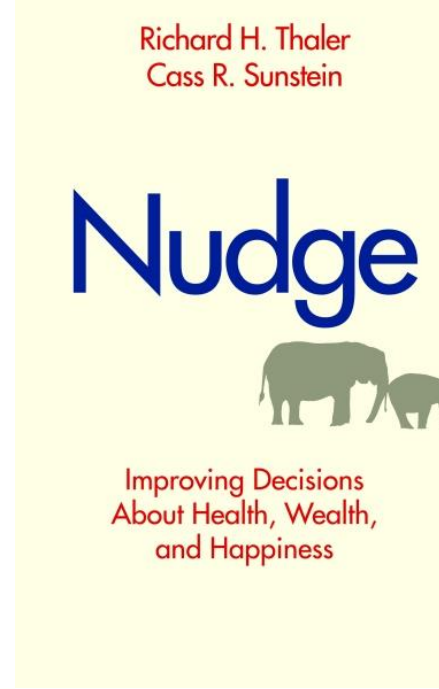- Group polarization
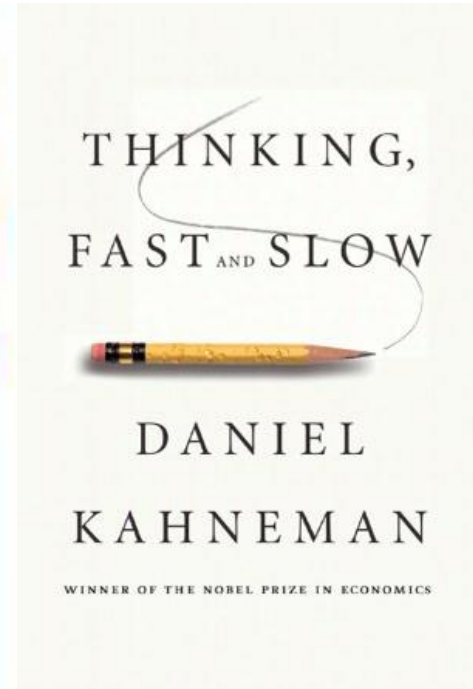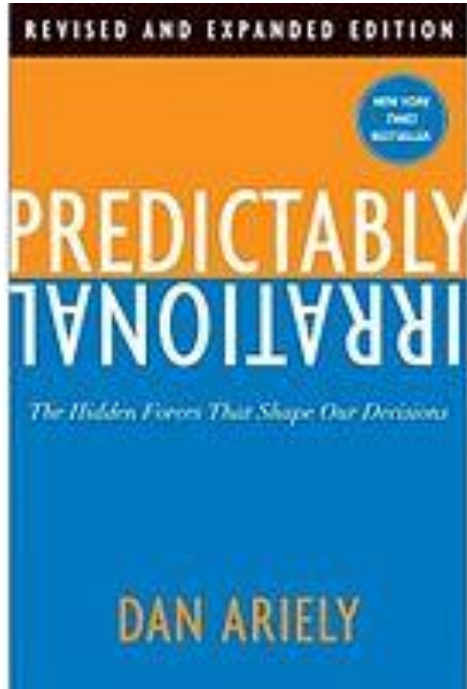
"Architectural" Errors

Daniel Kahneman

Amos Tversky

"Architectural" Errors

"Architectural" Errors

- "We use heuristics, or rules of thumb, that lead us to make predictable errors."
- Availability heuristic:
  - Familiarity: biases perceptions in terms of what we know ("If you can think of it, it must be important" – Wikipedia)
  - Salience: television vs print news, recency

  - "Imagine if <candidate> was President"

"Architectural" Errors

- Framing effects:
  - "Of those who have this procedure, 90 percent are alive after five years" vs "Of those who have this procedure, 10 percent are dead after five years"
  - Write down the last two digits of your SSN. How much would you pay for X?

"Architectural" Errors

- Representativeness heuristic:
  - Estimate quantities based on how representative it is in your own experience
  - If you've only met people from CT who were rich, when asked if CT people are rich you might say yes because of your experience, not because of the true numbers

"Architectural" Errors

- Conjunction errors:
  - *Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*

  Which is more probable?
  - Linda is a bank teller.
  - Linda is a bank teller and is active in the feminist movement.

"Architectural" Errors

- Egocentric bias:
  - We believe others are like us
    - What percentage of people like computers?
- Hindsight bias:
  - Our estimation of what we would have done knowing the outcome doesn't match what we would have done
    - Would the Giants win the Superbowl?

Amplifying "Architectural" Errors

- Groups that deliberate amplify the effects of these architectural errors
    - Amplify reliance on the representativeness heuristic
    - Increase framing effects
    - More conjunction errors

Amplifying "Architectural" Errors

- Groups are more likely than individuals to escalate commitment to a wrong path
  - Increases with great identification with the group

Amplifying "Architectural" Errors

- Groups lead to decreased use of the availability heuristic
- Groups lead to decreased use of the egocentric bias
- Groups are less susceptible to hindsight bias

Hidden Profiles and Common Knowledge

- Hidden profiles: Information that is present across group members but that they do not find

- Common knowledge: Groups focus on the information that they share rather than the information that they don't

- People with the most common knowledge "usually have a disproportionate influence in discussion"

Hidden Profiles and
Common Knowledge

- "when key information is unshared, groups are more likely to select a bad option after discussion than would their individual members before discussion"

- Increases with group size

Hidden Profiles and Common Knowledge

- Low-status individuals are less likely to provide hidden information

- Group diversity impacts group deliberation

Cascades and Polarization

- Informational cascades:
  - Imagine polling people one by one
    - Person A: Answers X
    - Person B: Answers X
    - Everyone thereafter now faces an X bias, even if the two X's were random chance.

Cascades and Polarization

- Reputational cascades:
  - Imagine polling people one by one
    - Person A: Answers X
    - Person B: Answers X
    - You might not want to risk your reputation to disagree with them.
    - This cascades onward

Group Polarization

- Members of a deliberating group typically end up in a more extreme position in line with their tendencies before deliberation began

- Heightened with a sense of shared group identity

- In-group members have more force than out-group members