8. Rust, N.C. and Mehrpour, V. (2020) Understanding image memorability. *Trends Cogn. Sci.* 24, 557–568

9. Ferber, M.A. and Brun, M. (2011) The gender gap in citations: does it persist? *Fem. Econ.* 17, 151–158

10. Torres, L. and Eliassi-Rad, T. *et al.* (2020) The why, how, and when of representations for complex systems. *arXiv* 2006, 02870

11. Gutiérrez y Muhs, G., ed (2012) *Presumed Incompetent: The Intersections of Race and Class for Women in Academia*, University Press of Colorado

12. Dworkin, J. *et al.* (2020) (In)citing action to realize an equitable future. *Neuron* 106, 890–894

13. Paludi, M.A. and Strayer, L.A. (1985) What's in an author's name? Differential evaluations of performance as a function of author's name. *Sex Roles* 12, 353–361

14. MacNell, L. *et al.* (2015) What's in a name: exposing gender bias in student ratings of teaching. *Innov. High. Educ.* 40, 291–303

15. Moss-Racusin, C.A. *et al.* (2012) Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16474–16479

Science & Society

Of Viruses, Vaccines, and Variability: Qualitative Meaning Matters

Valerie F. Reyna[1],*

Deaths from COVID-19 depend on millions of people understanding risk and translating this understanding into risk-reduction behaviors. Although numerical information about risk is helpful, numbers are surprisingly ambiguous, and there are predictable mismatches in risk perception between laypeople and experts. Hence, risk communication should convey the qualitative, contextualized meaning of risk.

Amidst the COVID-19 crisis, it is crucial to understand how people think about risk and how this determines their risk-reduction behaviors [1]. As in other public health problems, outcomes hinge on people's choices: whether to practice social distancing to prevent the spread of COVID-19, safer sex to prevent HIV/AIDS, or vaccination to prevent seasonal influenza. However, there is a fundamental mismatch between how most people think about risk and the assumptions experts make about actual and ideal human thinking. That is, most people think about risk in terms of qualitative meaning, called gist, as opposed to the precise details of risk information [2]. This mismatch produces predictable pitfalls in risk communication that are avoidable.

Why Numbers Are Ambiguous

The mismatch between gist and precise representations of risk goes beyond merely rounding off numbers, lumping rather than splitting, or innumeracy – the numerical equivalent of illiteracy [3]. To be sure, numeracy is a good thing. Popular numeracy tests ask respondents about probabilities and risks, such as questions about how to convert frequencies into probabilities, order different probabilities, and discriminate lower from higher risks. Other tests ask people to estimate the values displayed in a bar graph [4]. It is important to be able to read a graph and to know that a 0.10 probability of contracting COVID-19 is higher than a 0.01 probability. Every day during the pandemic, graphs and numbers hurl past the public.

However, numeracy is not sufficient to understand risk. In fact, numbers are ambiguous in the way that words are ambiguous, perhaps more so [5,6]. Suppose that a person hears that the number of deaths in the USA has surpassed 80 000, that the risk of transmission of COVID-19 is 2–3 times greater than that of the seasonal influenza, and that the mortality rate is about 3% of reported cases (Box 1). Decisions to act depend on the meaningful essence of this information. A simple linear transformation of numbers to categories does not capture the essence of risk. A nonlinear transformation of numbers does not suffice either. For example, a probability of 3% of rain would be low, but a probability of death of 3% from COVID-19 is high. Context matters for meaning.

Much research in the decision sciences has been devoted to demonstrating that context biases risky decisions, even making people who are risk-avoiding become risk-seeking just by changing how the same underlying facts are described [7]. These biases illustrate the human tendency to focus on changes relative to a reference point [8]. For example, a woman consulting the Breast Cancer Risk Assessment tool online (https://bcrisktool.cancer.gov/) is likely to be relieved to discover that her risk of cancer is below average because it is less than that of the population rate of about 13%, but how should she interpret these numbers? The numbers do not tell her the most important thing, namely, whether her risk is low or high. Her actions, whether to be screened more often than the average woman, and emotions, whether to feel calm or anxious, hinge on her interpretation of the gist of the risk: What does this information mean in context?

Meaning in context does not mirror literal reality. Typically, people do not think using what are called 'verbatim representations' of information. They think in fuzzy imprecise ways that interpret reality. For example, during a recent meeting I attended, public-health experts pointed out that those who test negative for a genetic mutation that increases breast-cancer risk technically do not have the same probability of developing breast cancer as members of the general population. But what is the gist of their risk? Testing negative does not mean that they have zero risk. Rather, their risk is less than the population average but remains in the same ballpark – the bottom line is that they could still develop cancer and need to take measures to reduce their risk (e.g., screening). For those who test

negative, the numbers change (and risk relative to before the test was given declines) but the gist of absolute risk stays about the same. Having a sense of relative and absolute risk can be important in different ways for different decisions [9].

Predictable Disconnects between Laypeople and Experts: When Knowledge Provides Context

Unfortunately, people cannot look up their individualized risk for COVID-19 using an online tool. As the average person looks around, he or she is likely to perceive little risk from COVID-19. After all, few people have died out of a vast number of people in the state where he or she lives. This ratio competence – the ability to understand that probabilities depend on the frequency of target events relative to a reference class of target and nontarget frequencies – is present early in life and in nonliterate cultures [10]. Thus, the perception of low personal risk is understandable and is likely to evoke resistance to risk-reduction measures such as social distancing, especially when they involve extreme limitations on economic activity and human interaction.

Although the risks of COVID-19 might seem low, background knowledge provides more than facts. It provides the context for interpreting the meaning of numbers such as 3% risk of mortality. For example, experts realize that

seasonal influenza often kills <0.1% of those who are infected, making COVID-19 more than ten times more lethal than the seasonal influenza, which kills 30 000 to >60 000 annually in the USA alone. Furthermore, greater lethality and transmissibility of COVID-19 relative to seasonal influenza combine to produce an exponential explosion in serious cases. When only 1% of a population is infected, the threshold at which an epidemic can be contained has already been crossed. A risk threshold is not an arbitrary cutpoint; it is the point at which risk changes qualitatively not just quantitatively. The significance of flattening the curve is that, without social distancing, the number of serious cases will hit a categorical boundary: exceeding the capacity of the healthcare system and causing many preventable deaths. Therefore, despite what seem like tiny numbers to laypeople initially, public-health experts perceive the risk of COVID-19 to be high.

Note that background knowledge, and scientific literacy broadly, allows members of the public to recognize what is plausible – what is likely to be true – as opposed to necessarily providing memorized truths that directly contradict incoming misinformation [11]. For example, one might accurately argue that the link between vaccines and autism has only been studied for a limited number of childhood vaccines. This argument was made by a

vaccination opponent; it is perfectly logical and even true. However, the question is whether such a link is plausible given current scientific knowledge.

Misinformation takes root in ignorance when the world does not make sense [12]. For example, the causes of autism, multiple sclerosis, and narcolepsy are unknown. Susceptibility is fostered by mistrust and suspicion of those perceived as powerful elites (the government, the rich, and researchers working in secret laboratories) and 'the other' (e.g., the 'Wuhan virus'). Bias can occur regardless of political persuasion [13]. Most important, misinformation is effective when it makes sense of the world and troubling events in it, when it offers a qualitative meaning that draws together pieces of reality and interprets them. This meaning might be woefully incomplete, but it is unlikely to be challenged if people do not seek out disconfirmatory tests and if they limit their contacts to like-minded others [14]. Reality is not infinitely reinterpretable, however, which creates opportunities to reach the public by communicating more than the facts, that is, conveying what the facts mean.

A Coda for Cognition

I have argued for the important role of cognitive science in understanding how people perceive risk and take risk-reduction actions in response to health threats, such as COVID-19. Emotion and social biases seem more relevant to risk perception than cognitive representations because of stereotypes about what cognition is. The stereotypical view of cognition is that it is cold, deliberative, and involves nothing more than educating people about rote facts. Perhaps this view is influenced by our reliance on the computer, and more recently machine learning, as metaphors for human cognition. When considering risk perceptions and responses related to such issues as HIV/AIDS, vaccination, and COVID-19, social, motivational, and emotional factors might seem paramount.

Certainly, all of these factors (along with many others such as culture, worldview, and experiences) matter in human responses to risk. But the interpretation of information and events surrounding risks – their qualitative meaning – is fundamental because the interpretation cues emotions,

motivations, and values. Qualitative does not imply noncomputational, because computations are merely tools for representing how people process information [15]. However, the meaningful imprecision of human cognition is not well captured yet in artificial intelligence. To a machine,

human beings can be defined as featherless bipeds without irony or bemusement. Humans chuckle. This definition is accurate in that it picks out the correct referents, but it omits the essence of what it means to be a human being.
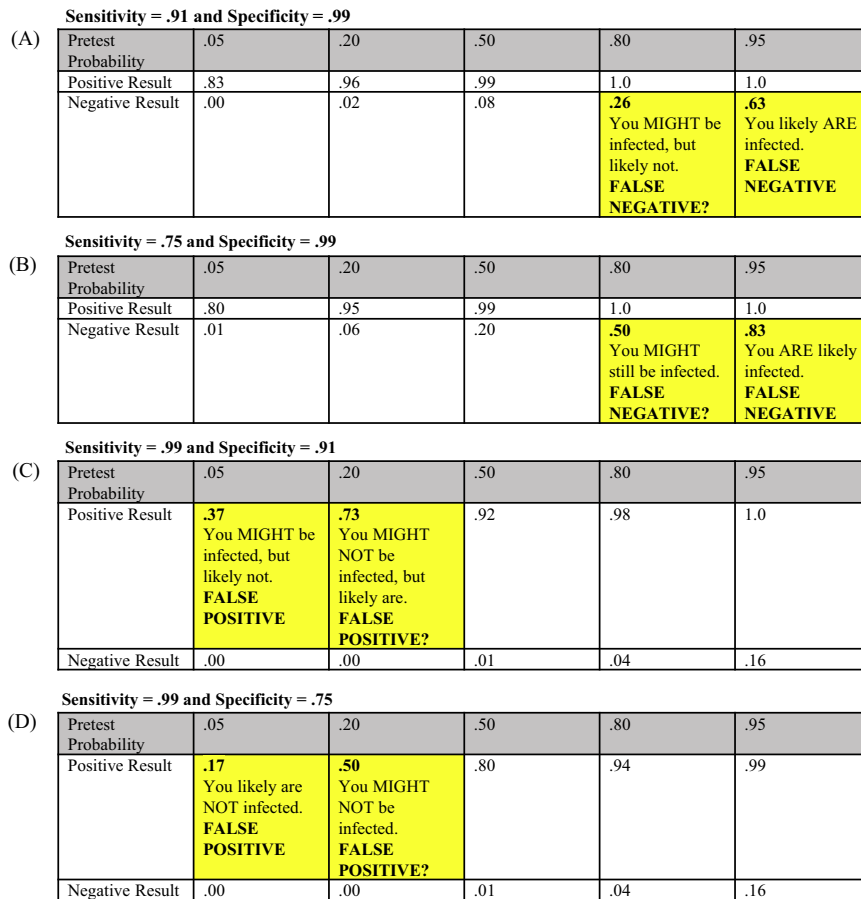
So, what can we do to better communicate risk? Begin with the end in mind: give people what they need to understand the qualitative, contextualized meaning of risk information. Figure 1 presents examples of how to combine pretest probabilities with COVID-19 testing to yield qualitative meanings. This approach has been applied to patients deciding among medications with serious side effects, teenagers making decisions about unprotected sex, and healthy people trying to figure out their genetic risk for cancer.

[1]Human Neuroscience Institute, MVR G331, Cornell University, Ithaca, NY 14853, USA

*Correspondence:
vr53@cornell.edu (V.F. Reyna).

https://doi.org/10.1016/j.tics.2020.05.015

(A) Sensitivity = .91 and Specificity = .99

Pretest Probability	.05	.20	.50	.80	.95
Positive Result	.83	.96	.99	1.0	1.0
Negative Result	.00	.02	.08	.26 You MIGHT be infected, but likely not. **FALSE NEGATIVE?**	.63 You likely ARE infected. **FALSE NEGATIVE**

(B) Sensitivity = .75 and Specificity = .99

Pretest Probability	.05	.20	.50	.80	.95
Positive Result	.80	.95	.99	1.0	1.0
Negative Result	.01	.06	.20	.50 You MIGHT still be infected. **FALSE NEGATIVE?**	.83 You ARE likely infected. **FALSE NEGATIVE**

(C) Sensitivity = .99 and Specificity = .91

Pretest Probability	.05	.20	.50	.80	.95
Positive Result	.37 You MIGHT be infected, but likely not. **FALSE POSITIVE**	.73 You MIGHT NOT be infected, but likely are. **FALSE POSITIVE?**	.92	.98	1.0
Negative Result	.00	.00	.01	.04	.16

(D) Sensitivity = .99 and Specificity = .75

Pretest Probability	.05	.20	.50	.80	.95
Positive Result	.17 You likely are NOT infected. **FALSE POSITIVE**	.50 You MIGHT NOT be infected. **FALSE POSITIVE?**	.80	.94	.99
Negative Result	.00	.00	.01	.04	.16

Trends in Cognitive Sciences

Figure 1. Illustrations of How Prior Probability and Test Accuracy Combine to Determine Probability Once a Test Result Is Known. Laypeople and physicians can be easily confused by the fact that results of a good diagnostic test might be the opposite of the truth: saying you do NOT have disease when you DO and vice versa. Sensitivity is the probability of a positive test result when someone has COVID-19 infection. Specificity is the probability of a negative test result when someone does NOT have COVID-19 infection. Example using the data presented in B: Of 100 people, if prior probability is 0.95, then 95 people are infected and five are not. Of the infected 95, 75% test positive (about 71). That means the remaining 24 people test negative. Since 99% of the five not infected test negative (about five out of five), almost all of the negative cases – 24 out of 29 (83%) – are actually infected. Bottom line for examples A and B: when sensitivity is lowish and priors are high, a lot of infected people test negative, so being negative does not mean much. By contrast, when specificity is less than sensitivity (examples C and D), the test can say you have the disease when you do not have it. Bottom line for C and D: when specificity is lowish and priors are low, a lot of people who are not infected test positive, so being positive does not mean much. Suppose you have a limited number of tests. Should you only test people who are hospitalized and likely to have the disease? If you have tests A or B, probably not, because a positive test is not very informative, and a negative test is misleading. However, if you have tests C or D, testing high-risk patients could be informative. Therefore, sensitivity, specificity, and prior-to-test probability all matter in surprising ways, not just the test result. For probability calculator, see http://araw.mede.uic.edu/cgi-bin/testcalc.pl

References

1. Fischhoff, B. *et al.*, eds (2012) *Communicating Risks and Benefits: an Evidence-Based User's Guide*, U.S. Department of Health and Human Services, Food and Drug Administration
2. Reyna, V.F. (2020) A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proc. Natl. Acad. Sci. U. S. A.* Published online April 20, 2020. https://doi.org/10.1073/pnas.1912441117
3. Peters, E. (2020) *Innumeracy in the Wild: Misunderstanding and Misusing Numbers*, Oxford University Press
4. Garcia-Retamero, R. and Cokely, E.T. (2013) Communicating health risks with visual aids. *Curr. Direct. Psychol. Sci.* 22, 392–399
5. Bransford, J.D. and Johnson, M.K. (1972) Contextual prerequisites for understanding: some investigations of comprehension and recall. *J. Verbal Learn. Verbal Behav.* 11, 717–726
6. Reyna, V.F. and Brust-Renck, P.G. (2020) How representations of number and numeracy predict decision paradoxes: a fuzzy-trace theory approach. *J. Behav. Decis. Making.* Published online April 1, 2020. https://doi.org/10.1002/bdm.2179
7. Kahneman, D. (2012) *Thinking, Fast and Slow*, Farrar, Straus and Giroux
8. Köszegi, B. and Rabin, M. (2007) Reference-dependent risk attitudes. *Am. Econ. Rev.* 97, 1047–1073
9. Fagerlin, A. *et al.* (2011) Helping patients decide: ten steps to better risk communication. *J. Natl. Cancer Inst.* 103, 1436–1443

10. Fontanari, L. *et al.* (2014) Probabilistic cognition in two indigenous Mayan groups. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17075–17080

11. Reyna, V.F. (2012) Risk perception and communication in vaccination decisions: a fuzzy-trace theory approach. *Vaccine* 30, 3790–3797

12. Oliver, J. and Wood, T. (2014) Medical conspiracy theories and health behaviors in the United States. *JAMA Intern. Med.* 174, 817–818

13. Ditto, P.H. *et al.* (2019) At least bias is bipartisan: a meta-analytic comparison of partisan bias in liberals and conservatives. *Perspect. Psychol. Sci.* 14, 273–291

14. Sloman, S. and Fernbach, P. (2017) *The Knowledge Illusion: Why We Never Think Alone*, Riverhead Press

15. Parsons, S. (1995) Refining reasoning in qualitative probabilistic networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (Besnard, P. and Hanks, S., eds), pp. 427–434, Morgan Kaufman

Science & Society

Online Developmental Science to Foster Innovation, Access, and Impact

Mark Sheskin,[1,2,13,*]
Kimberly Scott,[3,13]
Candice M. Mills,[4]
Elika Bergelson,[5,6]
Elizabeth Bonawitz,[7]
Elizabeth S. Spelke,[8] Li Fei-Fei,[9]
Frank C. Keil,[1]
Hyowon Gweon,[10]
Joshua B. Tenenbaum,[3]
Julian Jara-Ettinger,[1]
Karen E. Adolph,[11]
Marjorie Rhodes,[11]
Michael C. Frank,[10]
Samuel A. Mehr,[8,12] and
Laura Schulz[3]

We propose that developmental cognitive science should invest in an online CRADLE, a Collaboration for Reproducible and Distributed Large-Scale Experiments that crowdsources data from families participating on the internet. Here, we discuss how the field can work together to further expand and unify current prototypes for the benefit of researchers, science, and society.

The Need for Large-Scale, Shared Infrastructure for Developmental Science

The COVID-19 pandemic has disrupted society, including how we use the internet. Technology companies are recognizing advantages of permanent 'work from home' policies, the healthcare industry is re-evaluating its core infrastructure to support telemedicine, and schools are pushing the current limits of remote learning. Many of these developments are not new ideas, but rather overdue accelerations and extensions of existing trends. Paralleling previous moves to online research in other areas of cognitive science (e.g., adult surveys on Amazon Mechanical Turk [1]), researchers are using the internet in multiple ways to support developmental science. Here, we argue that a unified platform for online developmental science would benefit both researchers and society.

Our call for large-scale coordinated action mirrors advances in other sciences. Consider the Hubble Space Telescope: although many meaningful discoveries can be and were made with individually owned telescopes, Hubble hosts large-scale collaborative research that would otherwise be impossible. Similarly, ImageNet [2], which crowdsourced a massive amount of image-labeling data, transformed machine object recognition.

For developmental science, we propose a unified, discipline-wide, online Collaboration for Reproducible and Distributed Large-Scale Experiments (CRADLE). This discipline-wide infrastructure would support 'big science' research that would otherwise be impossible, and would also be a transformative platform for many 'small science' studies to produce better and more reproducible data. An online CRADLE is advantageous even for studies that cannot be conducted online, such as those that rely on specialized physical spaces (e.g., visual cliff studies with infants), special recording equipment [e.g., electroencephalogram (EEG)], or interactions with physical objects (e.g., physical puzzles), in that it frees up resources for those in-person studies. Thus, it can increase the quality of all developmental science research.

Current Efforts Towards Large-Scale Infrastructure in Developmental Science

The internet is already used in several ways to support developmental science. The Child Language Data Exchange System (CHILDES [3]) allows the reuse of high-value data sets by thousands of researchers across many countries. The Databrary Project (Databrary.org [4]) likewise makes video data globally accessible. Laboratories use the internet to closely coordinate their in-person data collection, leading to distributed, large-scale data collection (e.g., the ManyBabies Consortium [5]).

Recently, several groups established platforms to collect data with children online, including unmoderated studies that are available any time (Discoveries Online [6]; Lookit [7]; and themusiclab.org [8]), and moderated studies that are scheduled video sessions with a researcher (TheChildLab.com [9]). Potential advantages of unmoderated studies include rapid collection of large samples and high consistency of study delivery; moderated studies enable live interaction with researchers to maximize child engagement and science outreach, and allow the study of social and communicative phenomena that require live interaction.

As multiple projects develop in parallel, researchers have begun to discuss advantages of larger collaborations. A first step in this direction is a joint website