



Communicating risk in prenatal screening: the consequences of Bayesian misapprehension

Gorka Navarrete^{1*}, Rut Correia² and Dan Froimovitch³

¹ Laboratory of Cognitive and Social Neuroscience, Department of Psychology, Universidad Diego Portales, UDP-INECO Foundation Core on Neuroscience, Santiago, Chile

² Faculty of Education, Universidad Diego Portales, Santiago, Chile

³ Department of Physiology, University of Toronto, Toronto, ON, Canada

*Correspondence: gorkang@gmail.com

Edited by:

David R. Mandel, Defence Research and Development Canada, Canada

Reviewed by:

Miroslav Sirota, King's College London, UK

Simon John McNair, Leeds University Business School, UK

Keywords: Bayesian reasoning, prenatal screening, health policies, risk communication, massive screening

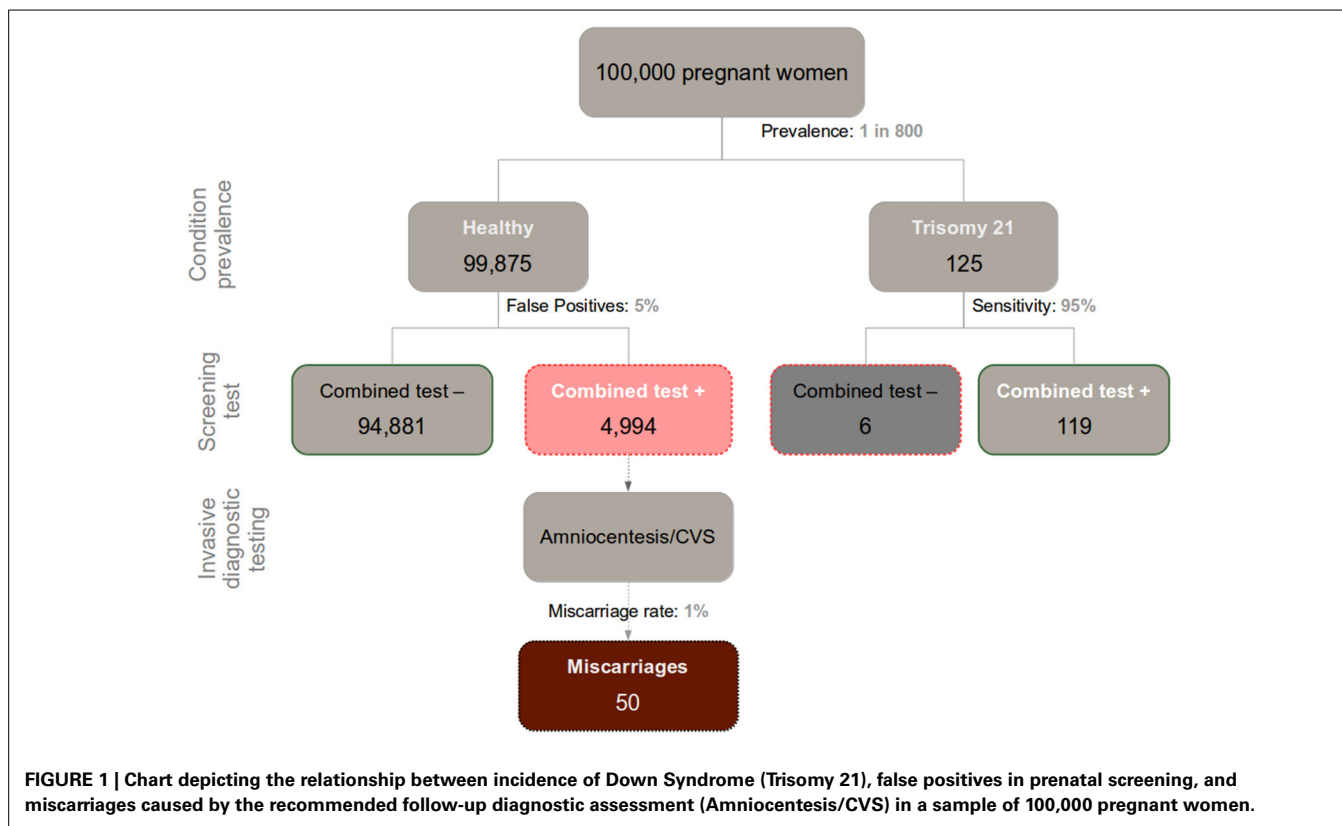
At some point during pregnancy women are typically encouraged to undergo a screening test in order to estimate the likelihood of fetal chromosomal aberrations. While timelines vary, the majority of pregnant women are screened within their first trimester (De Graaf et al., 2002). In the event of a positive test result, an invasive diagnostic assessment is usually recommended, namely amniocentesis or chorionic villus sampling (CVS). The combined test, widely considered to be the most feasible and effective screening procedure, involves an integrated assessment of: maternal age, fetal Nuchal Translucency (NT), maternal serum pregnancy-associated plasma protein A (PAPP-A), and free β human chorionic gonadotropin (β -hCG). This assay is most reliable when performed nearest to the 11th week of gestation (Malone et al., 2005), at which its detection rate and false positive rate for trisomy 21, in optimal conditions, are approximately 95 and 5%, respectively (Nicolaidis, 2004). A variety of competing screening techniques are available in the first trimester, and though we focus on the combined test in our example below, the point raised in this article applies to each of them.

A first-trimester screening assay carrying a relatively low false-positive rate might seem a reasonable option for women already considered to be at low risk—the vast majority of the pregnant population. Following such prenatal screening for trisomy 21, most women who test positive for high risk proceed with

invasive diagnostic testing. This decision to proceed with invasive testing is typically based on the presence of any evidence of increased risk brought to light by the precursory screening test (Nicolaidis, 2004). It is important to note, however, that the proportion of those who advance to invasive diagnostic testing is virtually identical to the false-positive rate of initial screening (Nicolaidis, 2004).

Applying trisomy 21 as an example (see **Figure 1** for a graphical representation of the numbers), the pregnant women who receive a false positive score in their first-trimester screening (~5%) would subsequently undergo a supplementary invasive diagnostic procedure, such as amniocentesis or CVS. This implies that out of every 100,000 pregnant women initially screened, roughly 5100 test positive, out of which ~5000 cases are actually false positives. The follow-up diagnostic tests are associated with serious procedure-related health risks, including a ~1% increased chance of miscarriage (see Mujezinovic and Alfirevic, 2007 for a systematic review; also, a recent nation-wide 11-year longitudinal study in Denmark established an increased chance of miscarriage of 1.4% and 1.9% linked to amniocentesis and CVS respectively, with CVS growing in its predominance worldwide; Tabor et al., 2009). Thus, at least 50 of the above ~5000 false-positive cases that involve normal fetuses ultimately result in diagnostic procedure-induced miscarriage. Of course with either a higher false-positive rate or a lower disease prevalence, those numbers worsen.

Discerning the trustworthiness of a given positive result in a screening test warrants calculating (typically from the information provided in the respective consent form) the test's positive predictive value (PPV; in this case the proportion of Down syndrome cases relative to the total amount of positive results). This requires knowledge of the base incidence rate of the congenital defect of interest, and the sensitivity and false-positive rate of the test. Computation and proper interpretation of this index, however, is often obscured by the complexity of Bayesian reasoning involved. This, among other factors, may underlie the well-known inadequacy of current procedures intended to achieve informed consent (Green et al., 2004). For 30-year-old pregnant women, the prevalence of Down syndrome is roughly 1 out of every 800 fetuses (Nicolaidis, 2004; this statistic varies with maternal age and time-point during pregnancy). In a sample of 100,000 pregnant women of the general population, therefore, around 125 of them would be expected to carry a fetus with the condition. Given the relatively high sensitivity of the screening assay (95% in optimal conditions), a majority of those fetuses are eventually correctly diagnosed with Down Syndrome (~119 out of 125). But when we merge this information with the said ~5000 false positives, we see that 119 positive results in the combined test faithfully reveal trisomy 21, out of a total 5113 (119 + 4994) positive results. Hence, the PPV of the combine test in a screening context nears 2% (119/5113). In other



words, there is a 2% chance of actually carrying a fetus with trisomy 21 after testing positive in a screening combined test. This information—essential to an educated decision on the matter—is usually overlooked by practitioners, and generally absent from medical consent forms.

In recent decades, our ineptitude for making sense of Bayesian information has been the subject of extensive study (for a review see Barbey and Sloman, 2007). It is widely recognized that humans struggle in dealing with Bayesian problems presented in terms of normalized probabilities (i.e., relative probabilities or percentages) or in cases of vague information structure (Barbey and Sloman, 2007). A substantial portion of the research on this topic has been done within the scope of medicine and epidemiology, wherein Bayesian inference pervades disease detection and characterization. It is well known that even medical practitioners struggle to interpret such information (Gigerenzer et al., 2007; but see Pighin et al., 2014 for a more optimistic outlook). The issue saliently manifests in the prevailing appeal of massive screening programs to the

general public, policy-makers, and physicians alike. This appeal—mainly due to the perceived advantages of early diagnosis—fails to be balanced by sufficient consideration of the high propensity for false alarms and over-diagnosis. The theoretic difficulties that most primary care physicians, for instance, seem to encounter with this type of information (e.g., cancer screening statistics) disposes them to a disproportionate veneration for the potential benefits of disease screening, as they drastically underrate the seriousness of relevant risks.

Gigerenzer et al. have advised on the pernicious use of massive screenings with respect to prostate cancer, HIV infection, etc. (Gigerenzer et al., 2007). False positives can be highly problematic in their ensuing psychosocial turmoil, and with respect to iatrogenic complications and economic costs associated with unnecessary clinical intervention. Moreover the problems, as we have seen above, don't stop at this. Medical knowledge ought to be conveyed lucidly, in a manner that facilitates informed decision-making, specifically accounting for the common cognitive challenges and inter-individual variation

observed in probability literacy (Johnson and Tubau, 2013; Lesage et al., 2013; Låg et al., 2014; Sirota et al., 2014a). With respect to clinical screening data, sufficient understanding of the numbers not only entails being in a position to competently evaluate pertinent risks; it further entails being enabled to recognize the possibility that even tests carrying low false-positive rates may simply be inadequate for detecting low-prevalence diseases, particularly in massive-screening settings.

There is growing convergence in cognitive psychology regarding the chief factors that mediate computation of Bayesian reasoning problems. Furthermore some practical improvements in the communication of statistical information have been proposed (while focus on evolutionary underpinnings of these issues appears to have taken a back seat in the literature (Barbey and Sloman, 2007; Navarrete and Santamaría, 2011)). With respect to understanding Bayesian problems, apart from intrinsic differences across individuals, in cognitive resources (Lesage et al., 2013; Låg et al., 2014; Sirota et al., 2014a) or numeracy skill (Hill and Brase, 2012; Johnson

and Tubau, 2013; Låg et al., 2014), several other factors that pertain to informational presentation *per se* have been deemed relevant to reasoning performance. These include (but are not limited to): problem structure (Barbey and Sloman, 2007; Lesage et al., 2013; Sirota et al., 2014a), the availability of a causal framework (Krynski and Tenenbaum, 2007), representational format (Hoffrage et al., 2002), and reference class (Fiedler et al., 2000; Lesage et al., 2013). Over and above intellectual aptitude, the very manner in which a problem's terms are conveyed to the subject is arguably imperative to the normative Bayesian response.

The above theoretical advancements have translated into numerous helpful strategies for representing and communicating Bayesian information. Regarding medical risk problems, if a subject is provided with the relevant information comprising the standard menu (i.e., hit rate, false positive rate and prevalence; Gigerenzer and Hoffrage, 1995), the most effective way known to facilitate reasoning is to ensure that the problem's set structure is entirely clarified to the subject (Barbey and Sloman, 2007). Natural frequencies (Gigerenzer and Hoffrage, 1995), or more generally, absolute reference classes (Fiedler, 2000; Lesage et al., 2013) are widely considered instrumental to this end. Another important factor, admittedly difficult to disentangle conceptually from the previous one, is computational complexity (Gigerenzer and Hoffrage, 1995; Barbey and Sloman, 2007). Reducing a subject's need to carry out computations (even those of simple arithmetic operations) can substantially enhance reasoning performance. Moreover the use of iconic and interactive representations has been shown to improve performance accuracy (Brase, 2009; Tsai et al., 2011; Micallef et al., 2012; Sirota et al., 2014b). Finally, an increasingly important area of research in this regard pertains to the development of training-programs designed to improve patients' and physicians' comprehension and computation of Bayesian problems (Sedlmeier and Gigerenzer, 2001; Sirota et al., 2014c).

There is a persistent need for advancing research concerning efficacious communication of Bayesian information, such that it can be comprehended by as many

individuals as possible—most urgently, those who intervene in health care decision making, such as clinicians and policy-makers. Wide-scale disease screenings hold both advantages and drawbacks (Gigerenzer et al., 2007), and a clear cognizance of their performance characteristics and the numbers underlying them is crucial to the state of public health and safety. At the moment, however, sufficient understanding of them is strikingly scarce, and with each passing year an unacceptable number of prospective parents are pressed to carry out a critical decision of potentially daunting consequences, without adequate knowledge of the important risks. And, of course, the quintessential challenges inherent to Bayesian reasoning are appreciable well beyond the domain of prenatal screening, posing egregious threats to the security and well-being of both the individual and the public.

ACKNOWLEDGMENTS

This research has been supported by the Semilla grants from the University Diego Portales (SEMILLA201418).

REFERENCES

- Barbey, A. K., and Sloman, S. A. (2007). Base-rate respect: from ecological rationality to dual processes. *Behav. Brain Sci.* 30, 241–254. doi: 10.1017/S0140525X07001653
- Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Appl. Cogn. Psychol.* 23, 369–381. doi: 10.1002/acp.1460
- De Graaf, I. M., Tijmstra, T., Bleker, O. P., and van Lith, J. M. M. (2002). Women's preference in Down syndrome screening. *Prenat. Diagn.* 22, 624–629. doi: 10.1002/pd.358
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* 107, 659–676. doi: 10.1037/0033-295X.107.4.659
- Fiedler, K., Brinkmann, B., Betsch, T., and Wild, B. (2000). A sampling approach to biases in conditional probability judgments: beyond base rate neglect and statistical format. *J. Exp. Psychol. Gen.* 129, 399–418. doi: 10.1037/0096-3445.129.3.399
- Gigerenzer, G., Gaissmaier, W., Kurz-milcke, E., Schwartz, L. M., Woloshin, S., and Dartmouth, T. (2007). Helping doctors and patients make sense of health statistics. *Psychol. Sci. Public Interest* 8, 53. doi: 10.1111/j.1539-6053.2008.00033.x
- Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704. doi: 10.1037/0033-295X.102.4.684
- Green, J. M., Hewison, J., Bekker, H. L., Bryant, L. D., and Cuckle, H. S. (2004). Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review. *Health Technol. Assess.* 8, 1–109.
- Hill, W. T., and Brase, G. L. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *Q. J. Exp. Psychol.* 65, 2343–2368. doi: 10.1080/17470218.2012.687004
- Hoffrage, U., Gigerenzer, G., Krauss, S., and Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition* 84, 343–352. doi: 10.1016/S0010-0277(02)00050-1
- Johnson, E. D., and Tubau, E. (2013). Words, numbers, and numeracy: diminishing individual differences in Bayesian reasoning. *Learn. Individ. Differ.* 28, 34–40. doi: 10.1016/j.lindif.2013.09.004
- Krynski, T. R., and Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *J. Exp. Psychol. Gen.* 136, 430–450. doi: 10.1037/0096-3445.136.3.430
- Låg, T., Bauger, L., Lindberg, M., and Friborg, O. (2014). The role of numeracy and intelligence in health-risk estimation and medical data interpretation. *J. Behav. Decis. Mak.* 27, 95–108. doi: 10.1002/bdm.1788
- Lesage, E., Navarrete, G., and De Neys, W. (2013). Evolutionary modules and Bayesian facilitation: the role of general cognitive resources. *Think. Reason.* 19, 27–53. doi: 10.1080/13546783.2012.713177
- Malone, F. D., Canick, J. A., Ball, R. H., Nyberg, D. A., Comstock, C. H., Bukowski, R., et al. (2005). First-trimester or second-trimester screening, or both, for Down's syndrome. *N. Engl. J. Med.* 353, 2001–2011. doi: 10.1056/NEJMoa043693
- Micallef, L., Dragicevic, P., and Fekete, J. (2012). Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE Trans. Vis. Comput. Graph.* 18, 2536–2545. doi: 10.1109/TVCG.2012.199
- Mujezinovic, F., and Alfirevic, Z. (2007). Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. *Obstet. Gynecol.* 110, 687–694. doi: 10.1097/01.AOG.0000278820.54029.e3
- Navarrete, G., and Santamaría, C. (2011). Ecological rationality and evolution: the mind really works that way? *Front. Psychol.* 2:251. doi: 10.3389/fpsyg.2011.00251
- Nicolaidis, K. H. (2004). *The 11–13+6 Weeks Scan*. London: Fetal Medicine Foundation.
- Pighin, S., Gonzalez, M., Savadori, L., and Giroto, V. (2014). Improving public interpretation of probabilistic test results: distributive evaluations. *Med. Decis. Mak.* doi: 10.1177/0272989X1436268. [Epub ahead of print].
- Sedlmeier, P., and Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *J. Exp. Psychol.* 130, 380–400. doi: 10.1037/0096-3445.130.3.380
- Sirota, M., Juanchich, M., and Hagmayer, Y. (2014a). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychon. Bull. Rev.* 21, 198–204. doi: 10.3758/s13423-013-0464-6
- Sirota, M., Kostovićová, L., and Juanchich, M. (2014b). The effect of iconicity of visual displays

- on statistical reasoning: evidence in favor of the null hypothesis. *Psychon. Bull. Rev.* 21, 961–968. doi: 10.3758/s13423-013-0555-4
- Sirota, M., Kostovičová, L., and Vallée-Tourangeau, F. (2014c). How to train your Bayesian. A problem-representation transfer rather than a format-representation shift explains training effects. *Q. J. Exp. Psychol.* doi: 10.1080/17470218.2014.972420. [Epub ahead of print].
- Tabor, A., Vestergaard, C. H. F., and Lidegaard, Ø. (2009). Fetal loss rate after chorionic villus sampling and amniocentesis: an 11-year national registry study. *Ultrasound Obstet. Gynecol.* 34, 19–24. doi: 10.1002/uog.6377
- Tsai, J., Miller, S., and Kirlik, A. (2011). Interactive visualizations to improve Bayesian reasoning. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 55, 385–389. doi: 10.1177/1071181311551079
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 28 September 2014; accepted: 20 October 2014; published online: 06 November 2014.
- Citation: Navarrete G, Correia R and Froimovitch D (2014) Communicating risk in prenatal screening: the consequences of Bayesian misapprehension. *Front. Psychol.* 5:1272. doi: 10.3389/fpsyg.2014.01272
- This article was submitted to *Cognition*, a section of the journal *Frontiers in Psychology*.
- Copyright © 2014 Navarrete, Correia and Froimovitch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.