A Probabilistic Solution for the Mystery of Japanese Blood Type Prediction Accuracy

Hideo Hirose, Kyushu Institute of Technology, Japan

Asian Conference on Psychology & the Behavioral Sciences 2015
Official Conference Proceedings

Abstract
In Japan, a belief that there is a relationship between personality and ABO blood types seems to have subsided now. Statistical hypothesis test results deny that relationship for many times up to now. However, many people still like to talk about blood types in connection with personality. They believe in predicting the blood type of one person with accuracy over a probability of 0.25 which is the mean value in providing the correct answer under the condition that they select one of the blood types at random. Actually, the questionnaire results we took several times reveal us an intriguing phenomenon; in Japan, blood type A and O were more accurately predicted than the random selection probability of 0.25. In the questionnaires, personality questions were not included, and this approach is new. In this research, we focused on the human intuition ability in predicting the blood types. That is, we investigated whether the correct answer ratio of human intuition is superior to the probability of random selection. Based on the conditional probability and Bayes theories, we show that this kind of superficial result can be expected as a natural phenomenon. We have revealed that people implicitly use the underlying knowledge of relationship between personality and ABO blood types as imprinting, which shows us superficial human intuition ability.

Keywords: blood type, personality, new questionnaire, human intuition, Bayes theory, imprinting

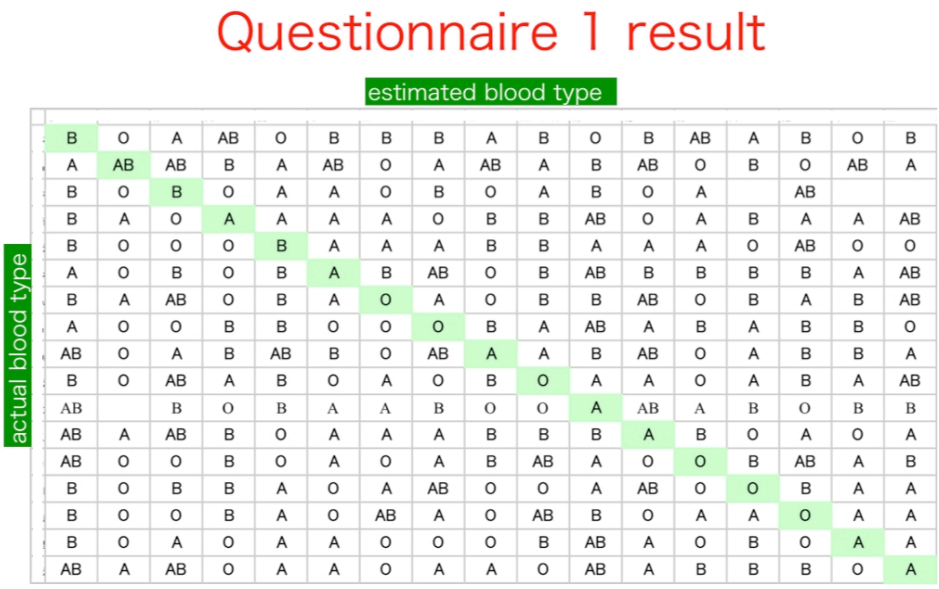
Introduction

In Japan, a belief that there is a relationship between personality and ABO blood types seems to have subsided now. Statistical hypothesis test results deny that relationship for many times up to now. A recent statistical research result also shows that such a relationship does not exist (Rogers & Glendon (2003), and Nawata, (2014)). Another approach rejecting the relationship by using the machine learning is seen in Hirose & Yukizane (2008), which is shown in appendix. However, many people still like to talk about blood types of persons. They may be interested in knowing personality of someone by an indirect method actually, and this underlying hidden information could be blood type. That is, implicitly, we are inclined to connect the blood types and personality.

On the other hand, many persons guess someone's blood type without knowing his/her personality. They believe in predicting the blood type of one person with accuracy over a probability of 0.25 which is the mean value in providing the correct answer under the condition that they select one of the blood types at random. It is very intriguing that blood type discussions in many cases are performed without personality. How can they guess other's blood types? This naïve question is one of the motivations of this paper.

Questionnaire 1

We first performed a new questionnaire in our research laboratory (B4, master students, and doctor students) because they know others to each other. The question is only one: "Guess someone's blood type." Figure 1 is a result of that. In the figure, row means the person who guesses someone's blood type, and column means the person whose blood type is estimated. Using successful and unsuccessful number of persons in estimation, a hypothesis test result is shown in Figure 2, where the randomness is not rejected. However, the number of samples was rather small.



B	O	A	AB	O	B	B	B	A	B	O	B	AB	A	B	O	B
A	AB	AB	B	A	AB	O	A	AB	A	B	AB	O	B	O	AB	A
B	O	B	O	A	A	O	B	O	A	B	O	A		AB		
B	A	O	A	A	A	A	O	B	B	AB	O	A	B	A	A	AB
B	O	O	O	B	A	A	A	B	B	A	A	A	O	AB	O	O
A	O	B	O	B	A	B	AB	O	B	AB	B	B	B	B	A	AB
B	A	AB	O	B	A	O	A	O	B	B	AB	O	B	A	B	AB
A	O	O	B	B	O	O	O	B	A	AB	A	B	A	B	B	O
AB	O	A	B	AB	B	O	AB	A	A	B	AB	O	A	B	B	A
B	O	AB	A	B	O	A	O	B	O	A	A	O	A	B	A	AB
AB		B	O	B	A	A	B	O	O	A	AB	A	B	O	B	B
AB	A	AB	B	O	A	A	A	B	B	B	A	B	O	A	O	A
AB	O	O	B	O	A	O	A	B	AB	A	O	O	B	AB	A	B
B	O	B	B	A	O	A	AB	O	O	A	AB	O	O	B	A	A
B	O	O	B	A	O	AB	A	O	AB	B	O	A	A	O	A	A
B	O	A	O	A	A	O	O	O	B	AB	A	O	B	O	A	A
AB	A	AB	O	A	A	O	A	A	O	AB	A	B	B	B	O	A

Figure 1: A result of questionnaire 1.

Hypothesis test

Hirose Labo.	A	B	O	AB	total
prevalence	6.8	3.4	5.1	1.7	17
observed	7	3	6	1	17
hits	33	16	24	0	73
vote	110	48	95	15	268
correct ratio	0.300	0.333	0.253	0.000	0.272
H0: p	0.250	0.250	0.250	0.250	
s.d. for ratio	0.0413	0.0625	0.0444	0.1118	0.0265
lower bound	0.167	0.125	0.161	0.026	0.197
upper bound	0.333	0.375	0.339	0.474	0.303
test	not reject	not reject	not reject		not reject

H0: p=0.25 not rejected

Figure 2: Hypothesis test to the result of questionnaire 1.

Then, we made the same questionnaire to a larger group, a class consisting of 80 sophomore students. Figure 3 shows the result of it. It is interesting that this time the correctly estimated ratio becomes around 0.3 in total, and this is included in the rejection region.

Hypothesis test 2

predicted	448	243	208	108	1007
sophomore	A	B	O	AB	total
prevalence	32	16	24	8	80
observed	32	23	16	9	80
hits	157	93	42	37	329
vote	331	291	200	185	1007
correct ratio	0.474	0.320	0.210	0.200	0.327
H0: p	0.250	0.250	0.250	0.250	
s.d. for ratio	0.0238	0.0254	0.0306	0.0318	0.0136
lower bound	0.202	0.199	0.189	0.186	0.223
upper bound	0.298	0.301	0.311	0.314	0.277
test	reject	reject	not reject	not reject	reject

H0: p=0.25 rejected or not rejected

Figure 3: Hypothesis test 2 to another result of questionnaire 1.

Figure 3 may be rearranged by a matrix form as shown in Figure 4, where areas indicate the difference among the blood types. It is clear that type A is dominantly occupied and that type AB is a minority group.

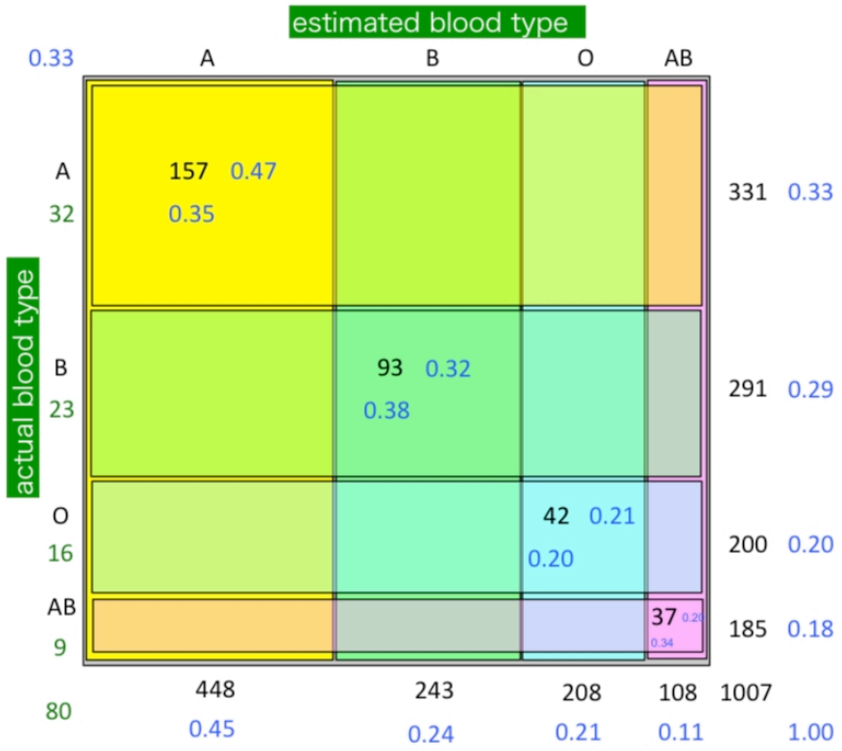


Figure 4: Matrix form of another result of questionnaire 1.

Using Bayes theorem, we can induce the estimation ratios as shown in Figure 5. These values are similar to the ratios computed from the observed values. From the Bayes perspective, we can estimated the blood types with accuracy about 40% for blood type A, about 20-30% for blood type B and O, and about 10% for blood type AB, when we used the observed values as prevalence values.

	A	B	O	AB	total		
A	157	73	68	33	331		0.329
B	119	93	60	19	291		0.289
O	95	44	42	19	200		0.199
AB	77	33	38	37	185		0.184
total	448	243	208	108	1007		1
	0.445	0.241	0.207	0.107	1		0.327

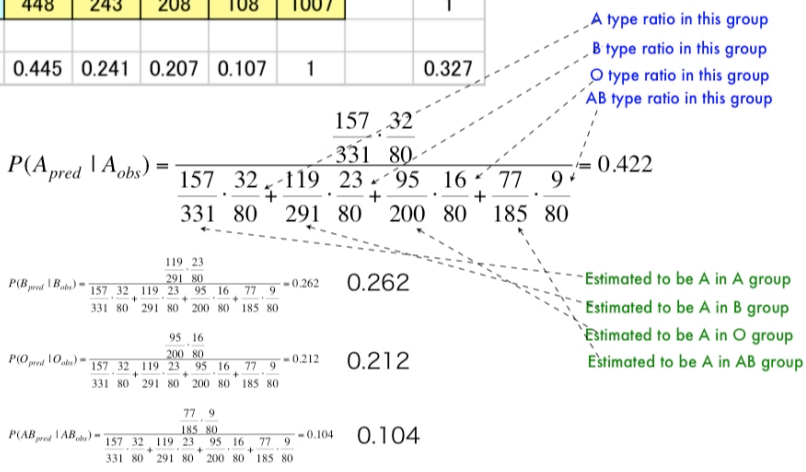


Figure 5: Bayes result to another result of questionnaire 1 (using observed ratios).

When we use the well-known common values as the prevalence values, we can derive the commonly used ratios as shown in Figure 6. That is, we can estimated the blood types with accuracy about 40% for blood type A, about 20% for blood type B, about 30% for blood type O, and about 10% for blood type AB. This means that the high accuracy in estimating blood type A seems to be merely a superficial phenomenon. We did not have high intuition abilities we might believe in. It can be explained just using the randomness. However, how can they know the prevalence values for A, B, O, and AB blood types? We consider next this naïve question.

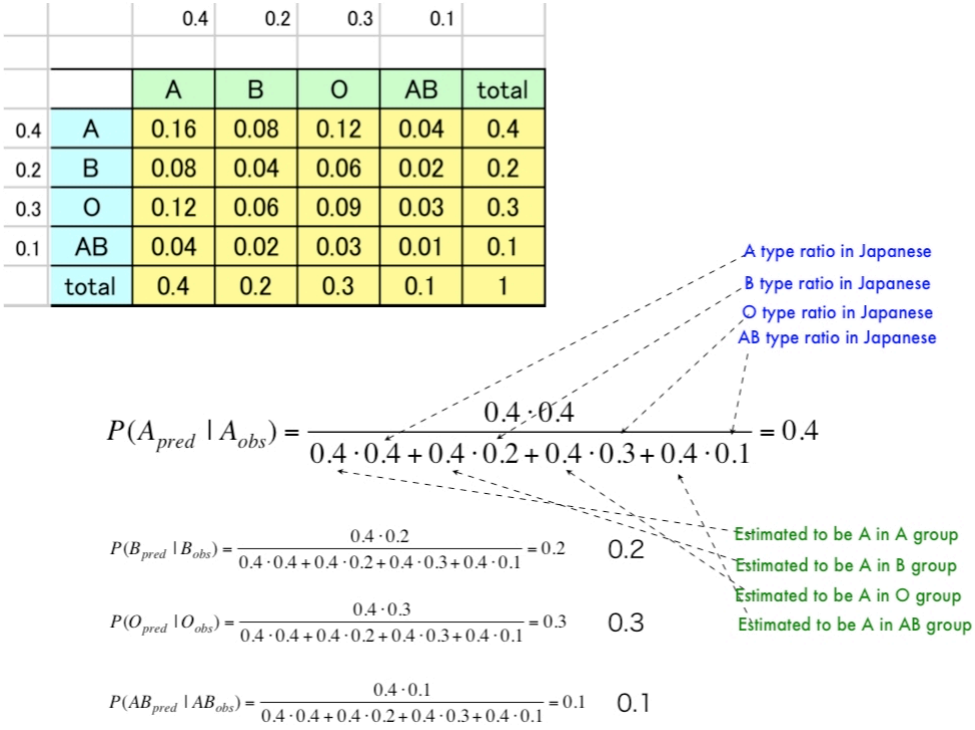


Figure 6: Bayes result to another result of questionnaire 1 (using prevalence ratios).

Questionnaire 2

Do we know the prevalence values for A, B, O, and AB blood types? If so, how can we know them? Then, we made a next questionnaire: “How many persons are there having characters in 10 persons shown below?” In many books explaining the relationship between blood types and personality, we can find typical connection example as shown in Table 1. This may be originated somewhere in the literature (e.g., Sakamoto, A., & Yamazaki, K. (2004)), and this information is commonly recognized in Japan.

Table 1: Commonly recognized personality in blood types A, B, O, and AB.

blood type	personality
A	scrupulous, steady, anxious mind
B	optimistic, own pace, freewheeling
O	broad-minded, sociable, unyielding
AB	unique, genius, dual personality

We may assume that the ratio of some blood type in total is due to the prevalence ratio of some characters in personality. For example, if we ask someone how many persons there are in 10 persons having characters, “scrupulous, steady, or anxious mind”. Then, he/she considers the percentage of persons having such characters. The characters above could be majority of Japanese persons because such characters are commonly seen in Japan. Actually, as Figure 7 shows the distribution to that question, they think there are many persons having such characters. The mean value is 3.2, and 50 percentile point is 3.1. More than three in ten persons are having such characters.

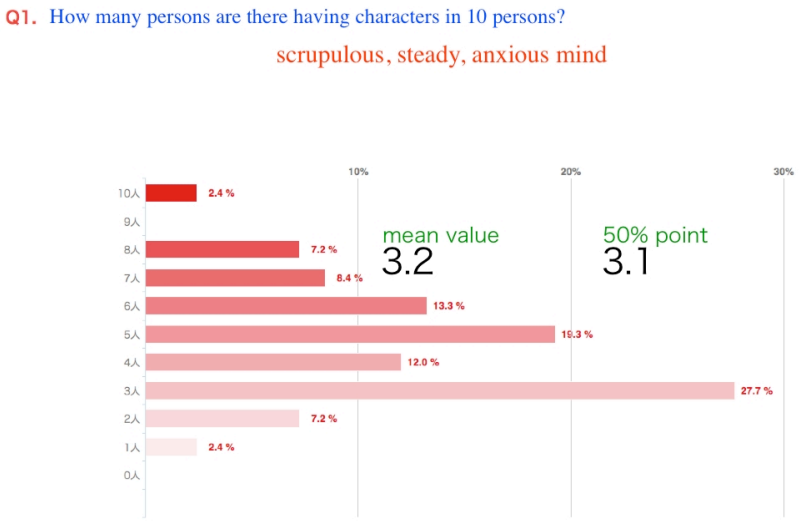


Figure 7: Result for the questionnaire: how many persons are there in 10 persons having characters, “scrupulous, steady, or anxious mind?”

Figures 8, 9 and 10 show the similar results for characters, “optimistic, own pace, freewheeling”, “broad-minded, sociable, unyielding”, and “unique, genius, dual personality”. The order of the ratios in ten persons to the questionnaire results reveals that the result in Figure 7 is the first, that in Figure 9 is the next, those in Figure 8 and 10 follow in succession. This order is consistent to the order of the blood type ratios: the first is A, next O, then B, and last AB.

The well-known prevalence ratios of Japanese blood types are 0.4 for type A, 0.2 for type B, 0.3 for type O, and 0.1 for type AB. The blood type AB is regarded as a minority group.

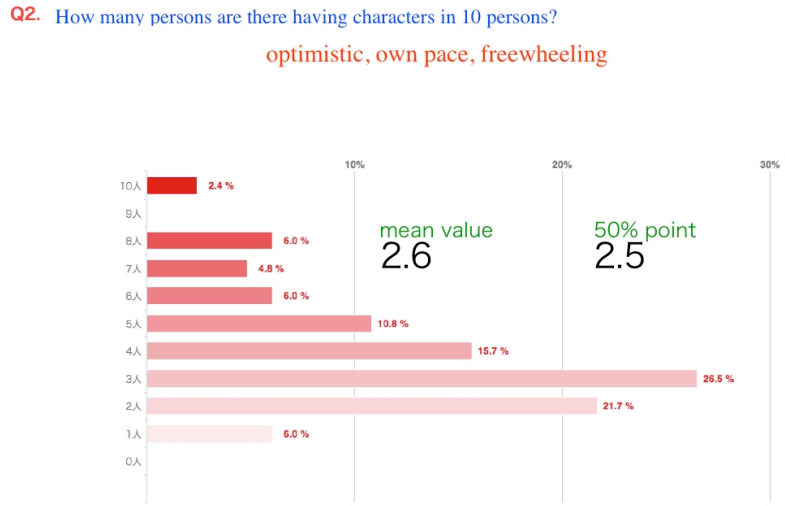


Figure 8: Result for the questionnaire: how many persons are there in 10 persons having characters, “optimistic, own pace, freewheeling?”

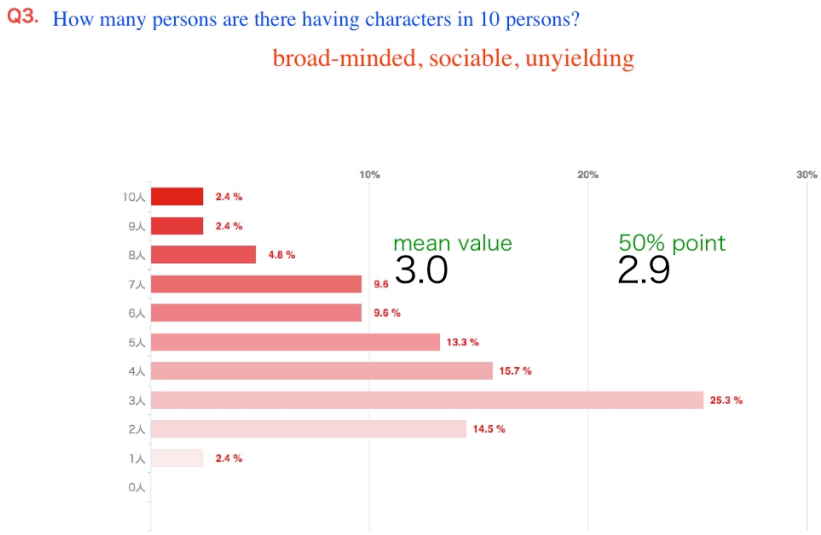


Figure 9: Result for the questionnaire: how many persons are there in 10 persons having characters, “broad-minded, sociable, unyielding?”

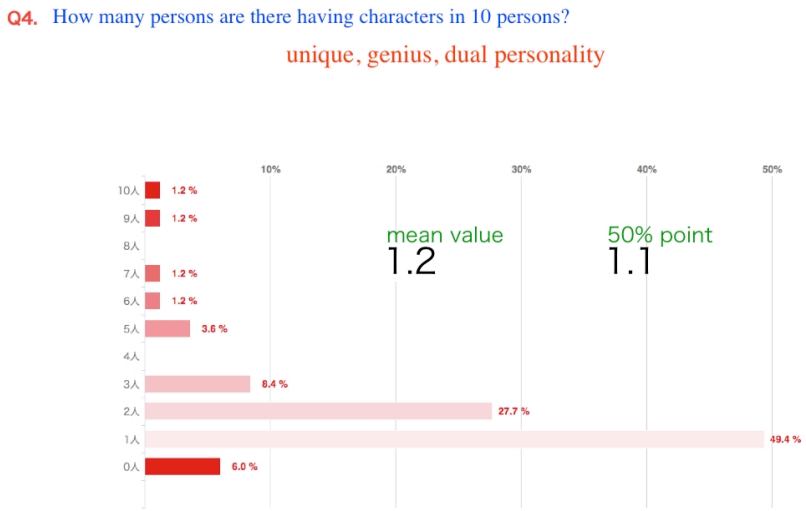


Figure 10: Result for the questionnaire: how many persons are there in 10 persons having characters, “unique, genius, dual personality?”

Figure 11 shows the box plot corresponding to Figures 7, 8, 9, and 10 although some transformation is performed. That is, for one person, total ratio summing up from four ratios to each blood type becomes one. In the figure, we can see some relationship between the blood types and the guessing numbers of persons having some personalities. Figure 12 shows the correlation between the ratios of actual prevalence one and those of the questionnaire results in mean values shown in Figure 11. The correlation coefficient value is 0.92. In anyway, the guessing numbers of persons having some personalities are consistent with the actual prevalence ratios of blood types.

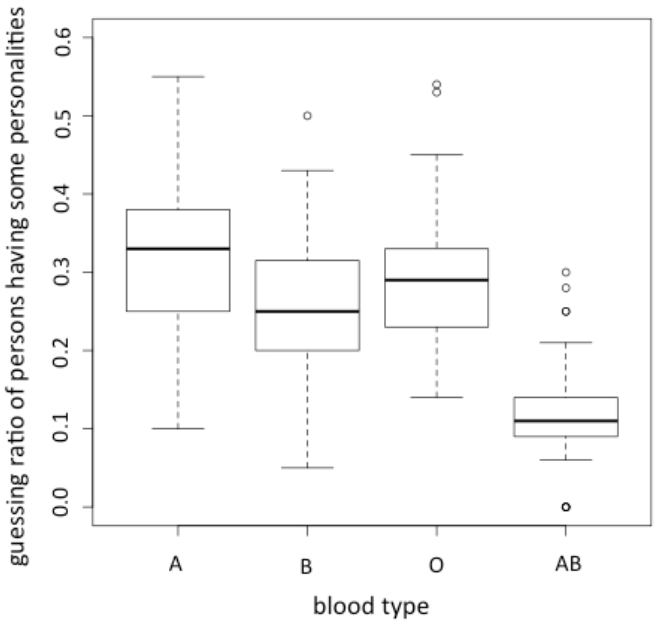


Figure 11: Correlation between the ratios of actual prevalence one and those of the questionnaire results shown in Figure 7, 8, 9, and 10.

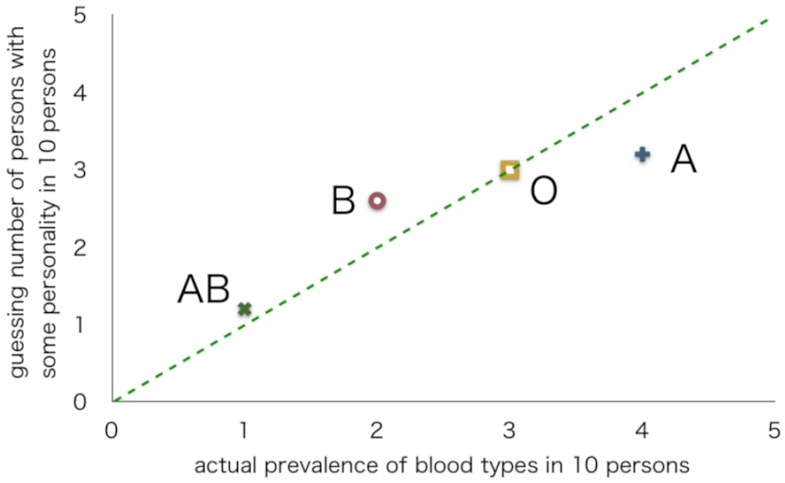


Figure 12: Correlation between the number of the actual prevalence and that of the questionnaire results in mean values shown in Figure 11.

Therefore, we could construct a hypothesis that many people guess someone's blood type via the hidden personalities. One can estimate the percentages of typical human characters, and these percentages happen to coincide with those of blood types. However, we cannot simply assume that there is a relationship between human characters and the blood types. Still, we can explain the accuracy of the estimated values even if people do not know the blood types prevalence. Without knowing the actual prevalence ratios of blood types, we can guess someone's blood type by with the aid of personality ratios commonly pervaded. This is why we can guess others' blood types more accurately (e.g., 0.3) than randomly estimated value (i.e., 0.25).

Discussion

Many researchers investigate the relationship between blood types and personality by using questionnaires in which blood types and personality are included altogether. This paper does not deal with the information of personality in questionnaires directly; it deals with only the blood type questionnaire: one can estimate someone's blood type. This is a novel point of the paper. If no information about personality is provided, the expected accuracy to correct answers seems to be 0.25 by randomness. However, a higher accuracy, e.g., 0.3, is obtained by our questionnaires. It seems that people first guess others' personality, and then they connect blood types and personality with their own information, and finally they guess blood types. If this is true, we have to assume that there may be some information about the relationship between the blood types and personality. Where do they get such the information? Are they medical or psychological references? We think that a source of the information could be imprinting developed over a long period of time through the literature. However, we know such the information is unreliable nowadays.

The accuracy to correct answers obtained by our questionnaire seems to be one that we know the prevalence ratios of blood types to the total. Therefore, we next tried to find the prevalence ratios. To find the hidden information between blood types and personality, we performed a different questionnaire: how many persons there are having such and such personality in ten persons.

If the ratios of personality people replied are consistent with the actual prevalence ratios, then we can estimate the blood types with a higher accuracy without the knowledge of the relationship between the blood types and personality. Only the ratios have critical importance. No meaning in the knowledge. This is our conjecture.

Conclusion

In Japan, a belief that there is a relationship between personality and ABO blood types seems to have subsided now, and statistical hypothesis test results deny that relationship for many times up to now. However, many people still like to talk about blood types in connection with personality. In the questionnaires, blood types and personality were included altogether. However, in this paper, personality questions were not included, and this approach is new. In this research, we focused on the human intuition ability in predicting the blood types. That is, we investigated whether the correct answer ratio of human intuition is superior to the probability of random selection. Based on the conditional probability and Bayes theories, we show that this kind of superficial result can be expected as a natural phenomenon. We have revealed that people implicitly use the underlying knowledge of relationship between personality and ABO blood types as imprinting, which shows us superficial human intuition ability. A subsequent questionnaire regarding the prevalence ratios of some personalities assisted the conjecture that only the prevalence ratios make sense but the relationship knowledge between the blood types and personality.

Appendix

As commonly used questionnaires, we also obtained two sets of database for the relationship between the blood types and personalities of as shown in Figure 13. One set is for sophomore class of fiscal year 2005, and the other is of 2006. A commonly used method in classification is the decision tree as shown in the figure.

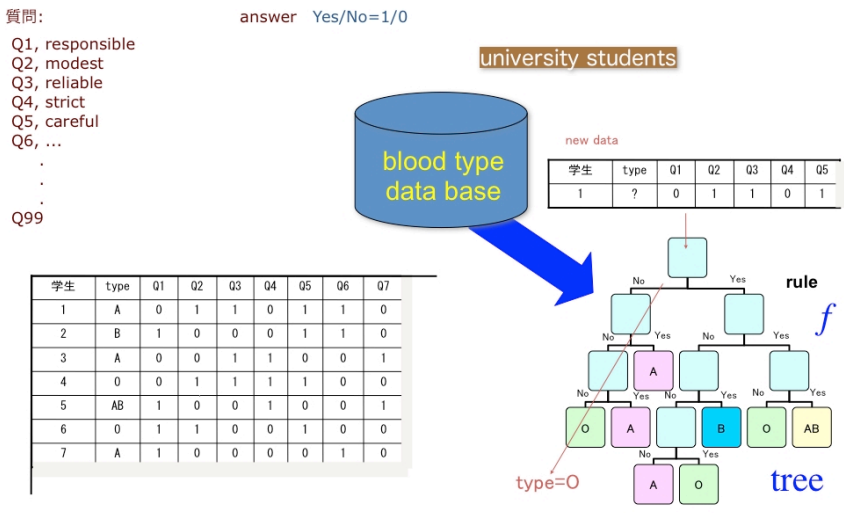


Figure 13: Questionnaire database and classification by using the tree.

We can construct the tree model by using the data set of 2005 and apply the model to the same data set, as shown in Figure 14. In such an application, about 60% accuracy for correct answers is obtained.

However, if we apply the model to the 2006 data set, the accuracy drastically reduced to 40%, which is commonly seen in other cases. This accuracy is also obtained if all the final leaves show the same class: blood type A is always estimated because A is the most major group. The decision tree automatically constructs such the model, and actually it does not work at all.

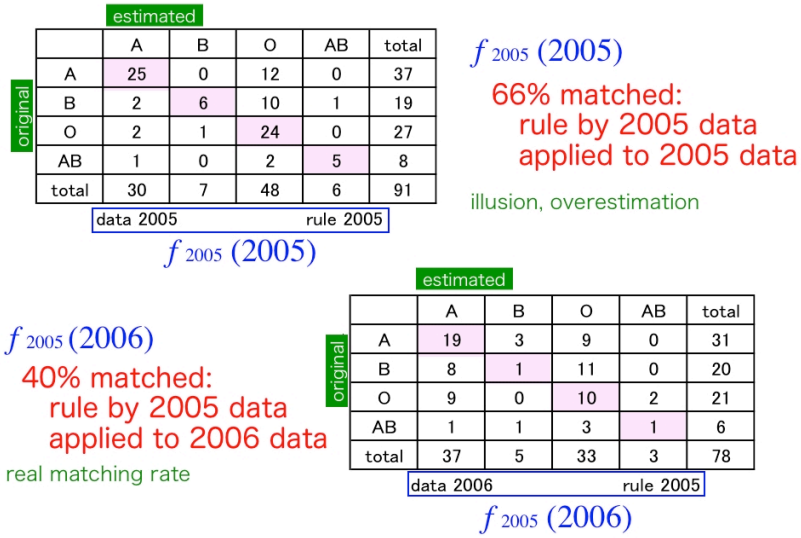


Figure 14: Superficial and actual classification accuracies.

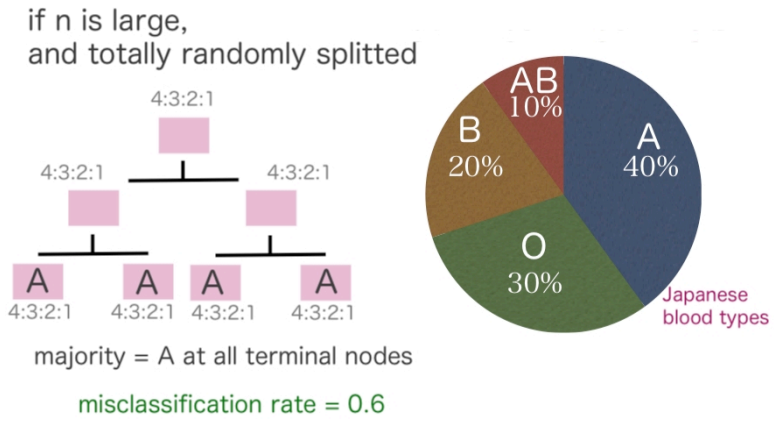


Figure 15: Japanese blood type prevalence ratios and the decision tree classification.

If we are in English-speaking world, the blood type prevalence ratios are seen in Figure 16. Then, we will no longer be able to estimate the blood types by using the same questionnaire 2 results.

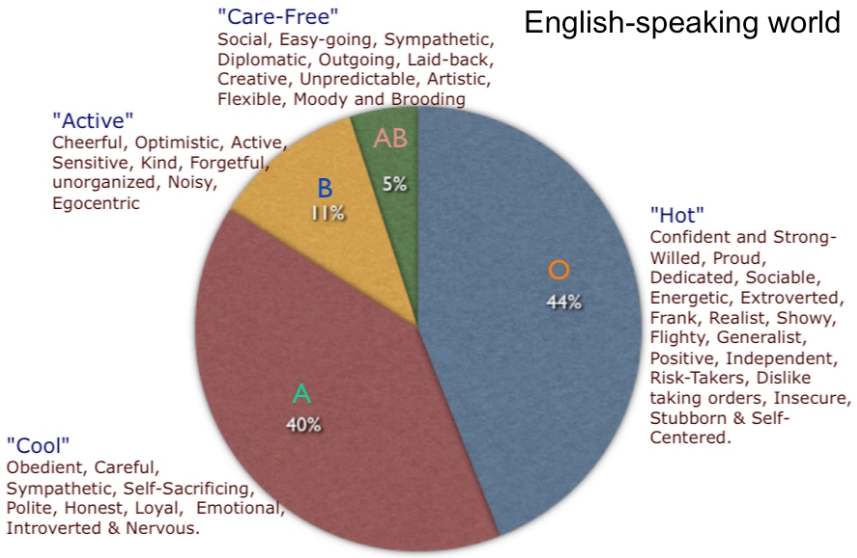


Figure 16: Blood type prevalence ratios in English-speaking world.

References

Cameron, D. (2001). *Working with spoken discourse*. London: SAGE.

Coates, J. (2007). Talk in a play frame: More on laughter and intimacy. *Journal of Pragmatics, 39*, 29-49.

Drew P., & Heritage J. (1992). Analysing talk at work: An introduction. In P. Drew, & J. Heritage (Eds.), *Talk at work* (pp. 3-65). Cambridge: Cambridge University Press.

Hirose, H. and Yukizane, T. (2008). The accuracy of the trade-off curve in the bump hunting, *7th Hawaii International Conference on Statistics, Mathematics and Related Fields*, Hawaii.

Nawata, K. (2014). No relationship between blood type and personality: Evidence from large-scale surveys in Japan and the US, *The Japanese Journal of Psychology*, **85**, 148-156.

Rogers, M., & Glendon, A. I. (2003). Blood type and personality. Personality and Individual Differences, 34,
1099–1112.

Sakamoto, A., & Yamazaki, K. (2004). Blood-typical personality stereotypes and self-fulfilling prophecy: A natural experiment with time-series data of 1978-1988. In Y. Kashima, Y. Endo, E. S. Kashima, C. Leung, & J. McClure (Eds.), Progress in Asian Social Psychology, Vol. 4. Seoul, Korea: Kyoyook-Kwahak-Sa. 239–262.

Contact email: hirose@ces.kyutech.ac.jp