LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INFORMED CHOICES, INCLUSIVE VOICES: EPISTEMIC JOURNEYS IN DEMOCRATIC DECISION MAKING

THE CONDORCET JURY THEOREM

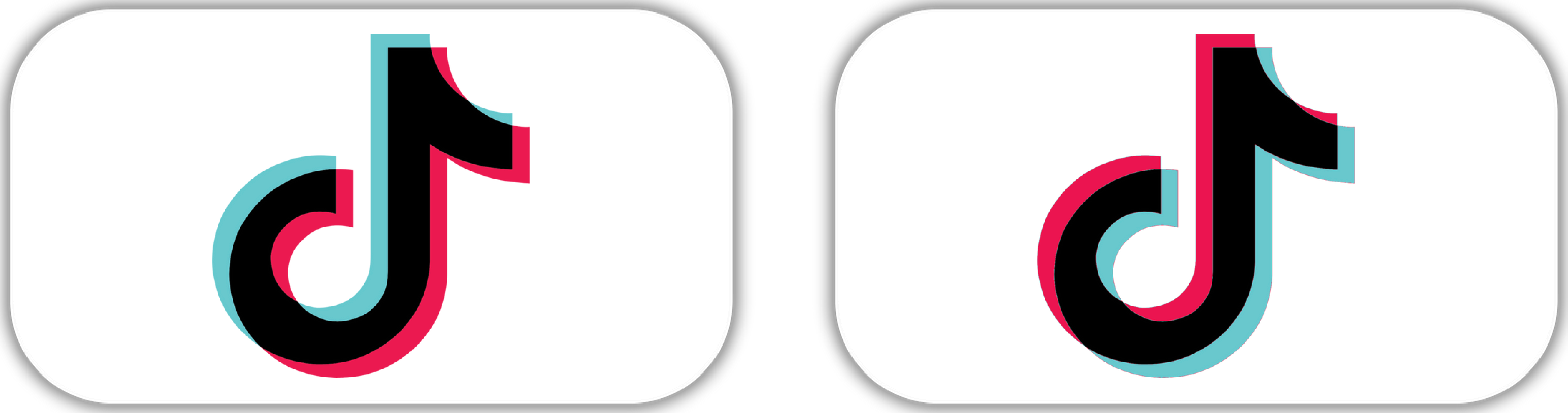
HOW GROUPS CAN BE SMART. IN VERY SPECIAL CONDITIONS...

Adrian Haret
a.haret@lmu.de

April 22, 2024

Let’s play a little game: try to guess the correct version of the logo.

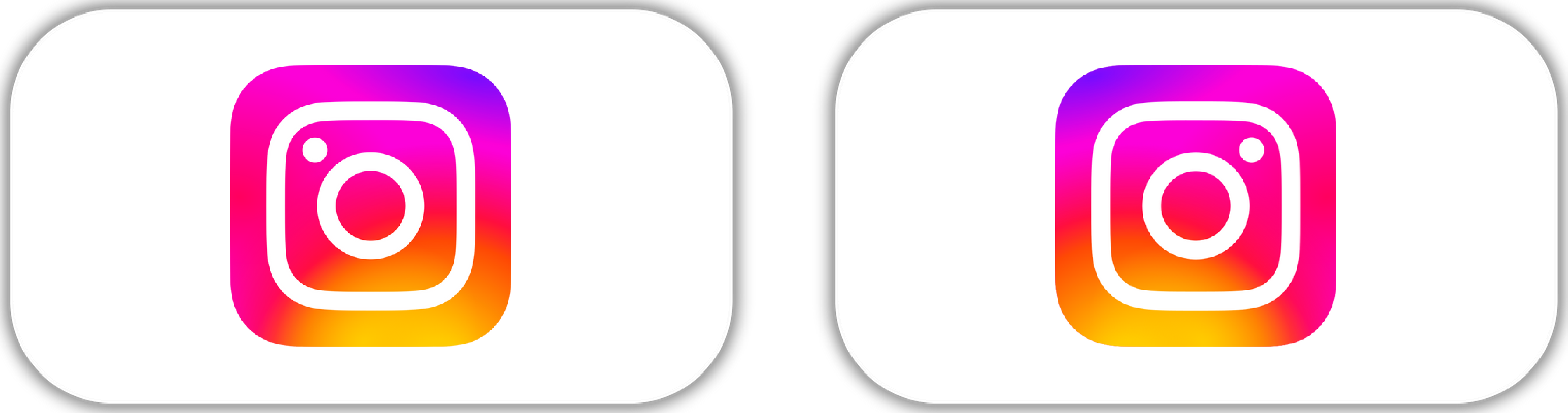
Keep track of your score!

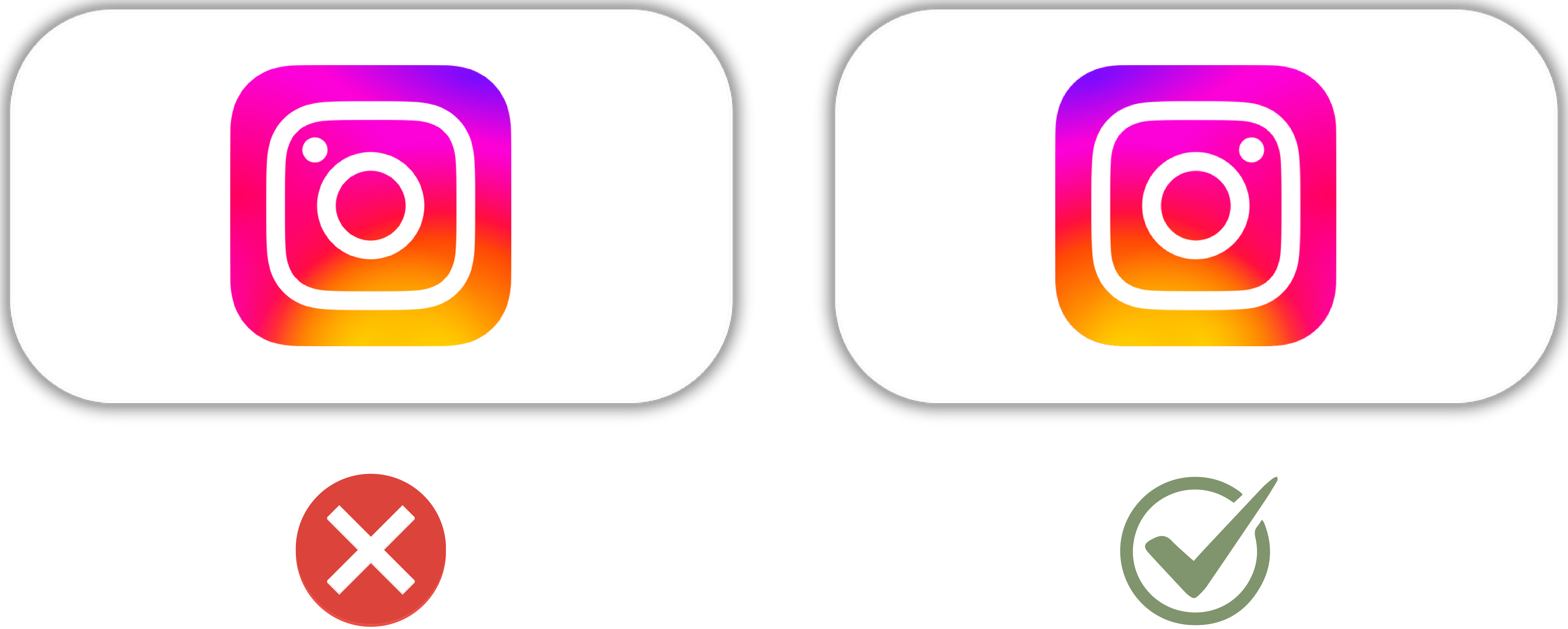


Google

Google

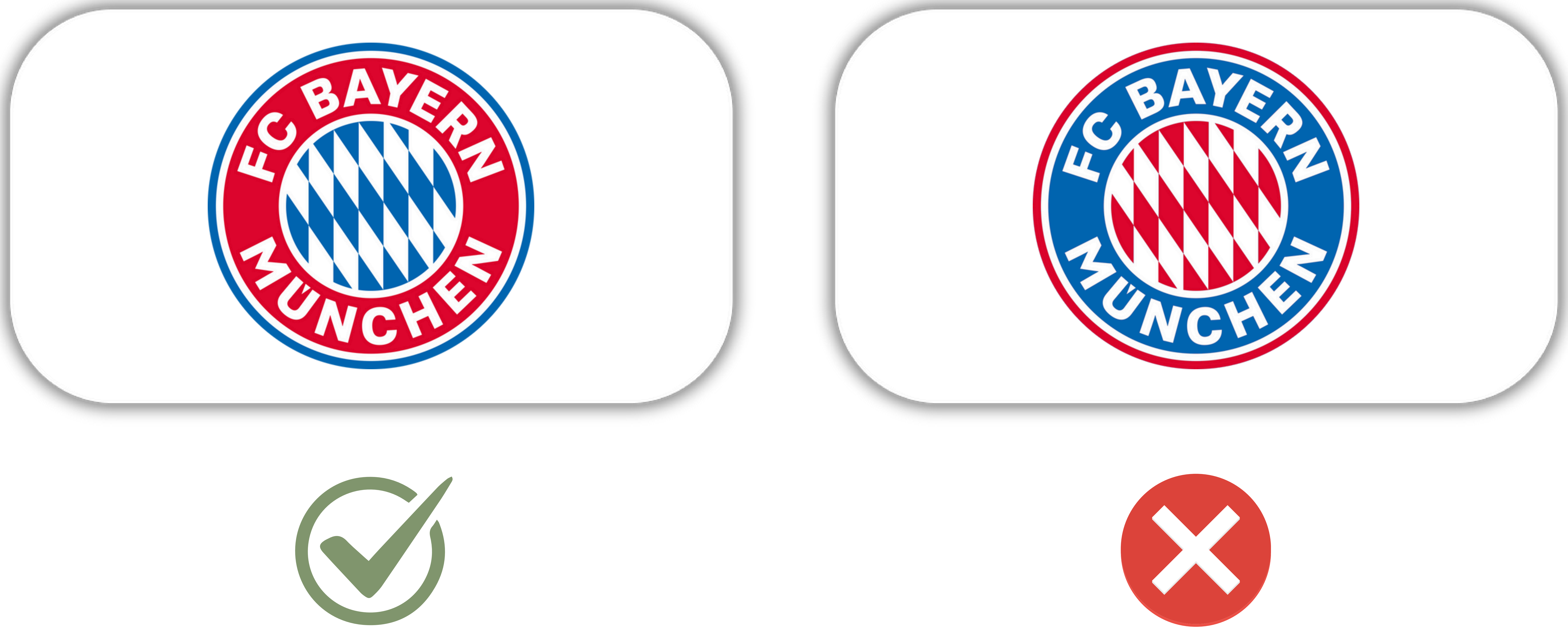

Google
Google

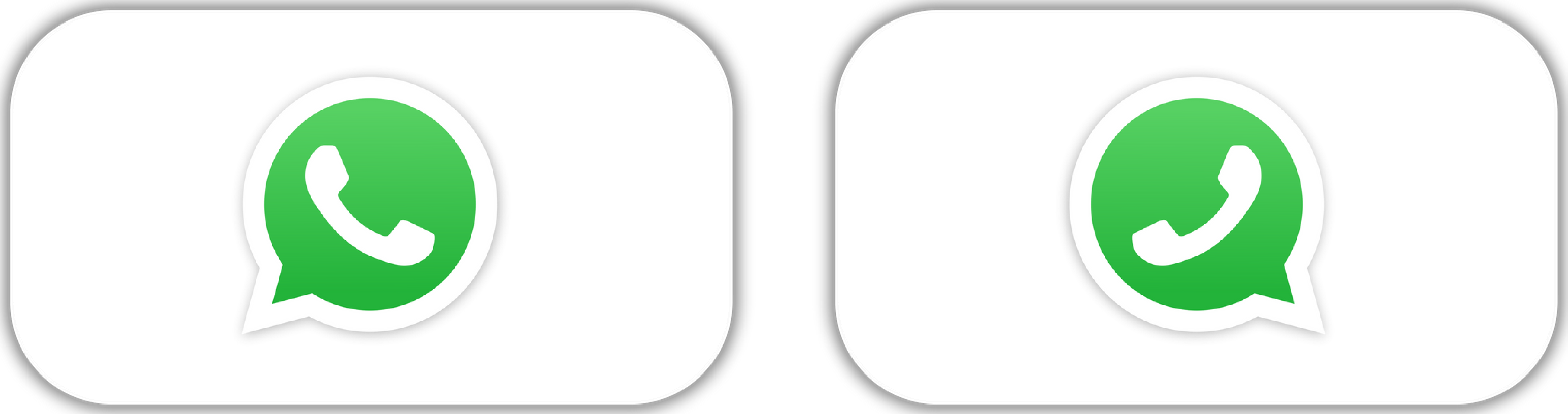


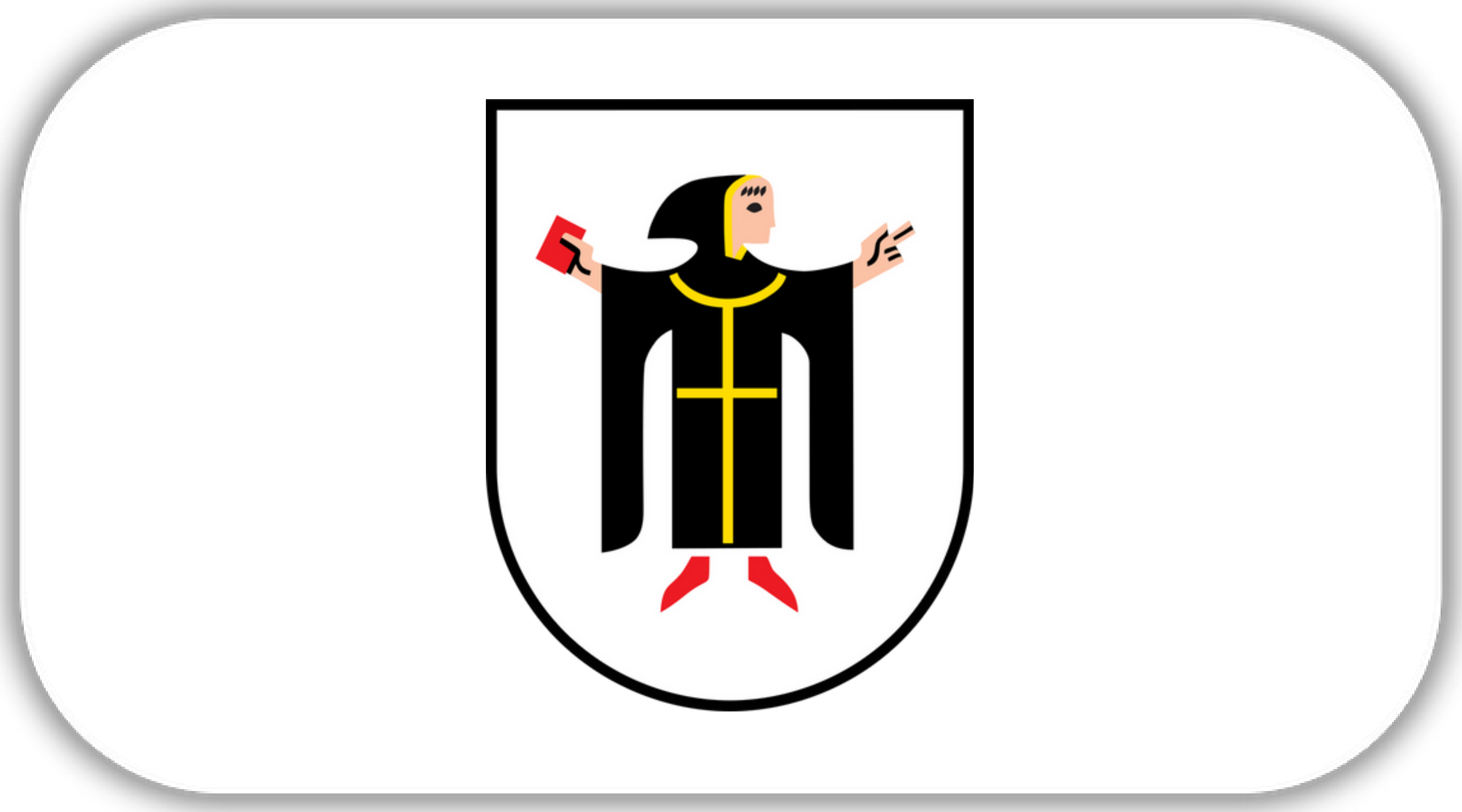


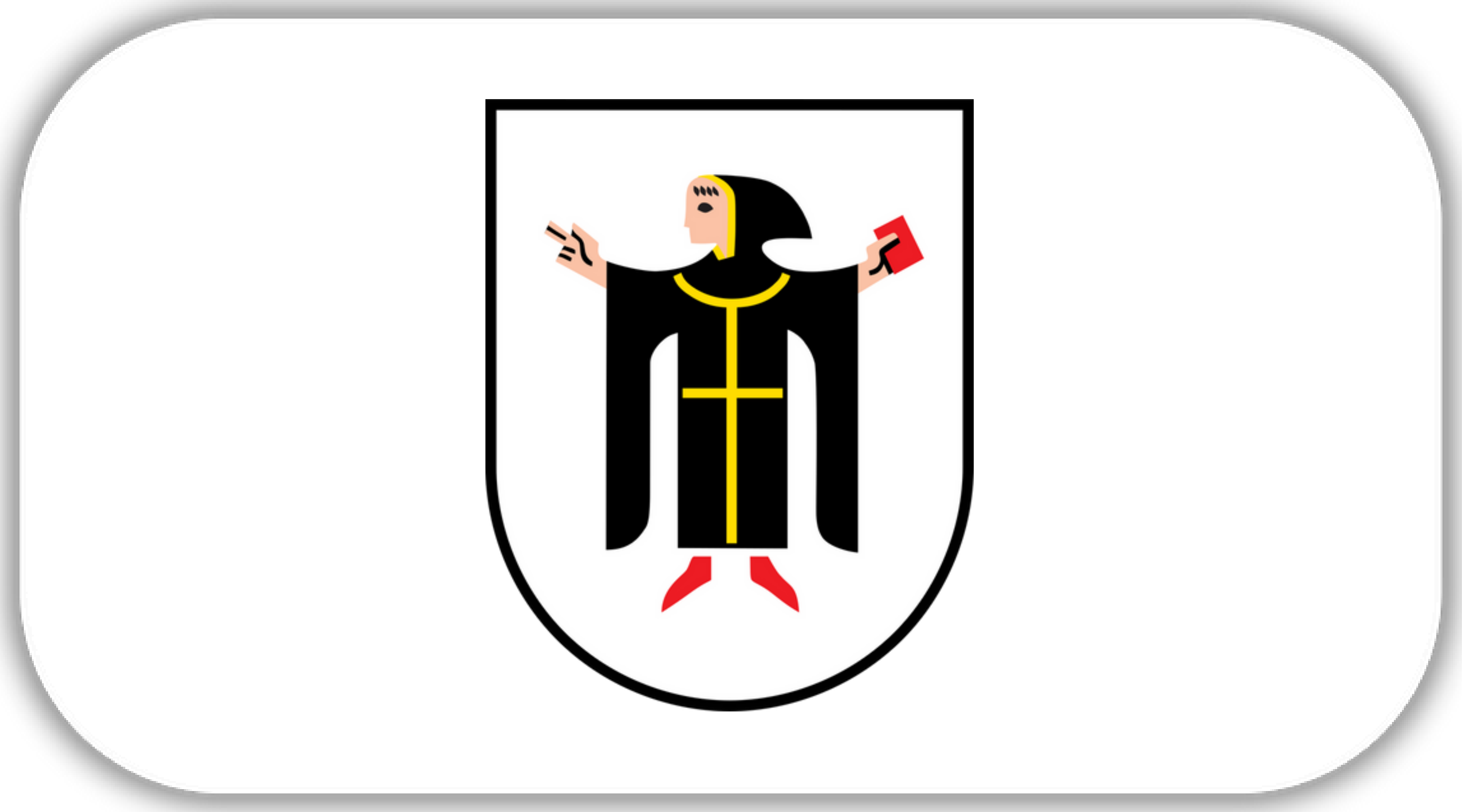
FC BAYERN MÜNCHEN

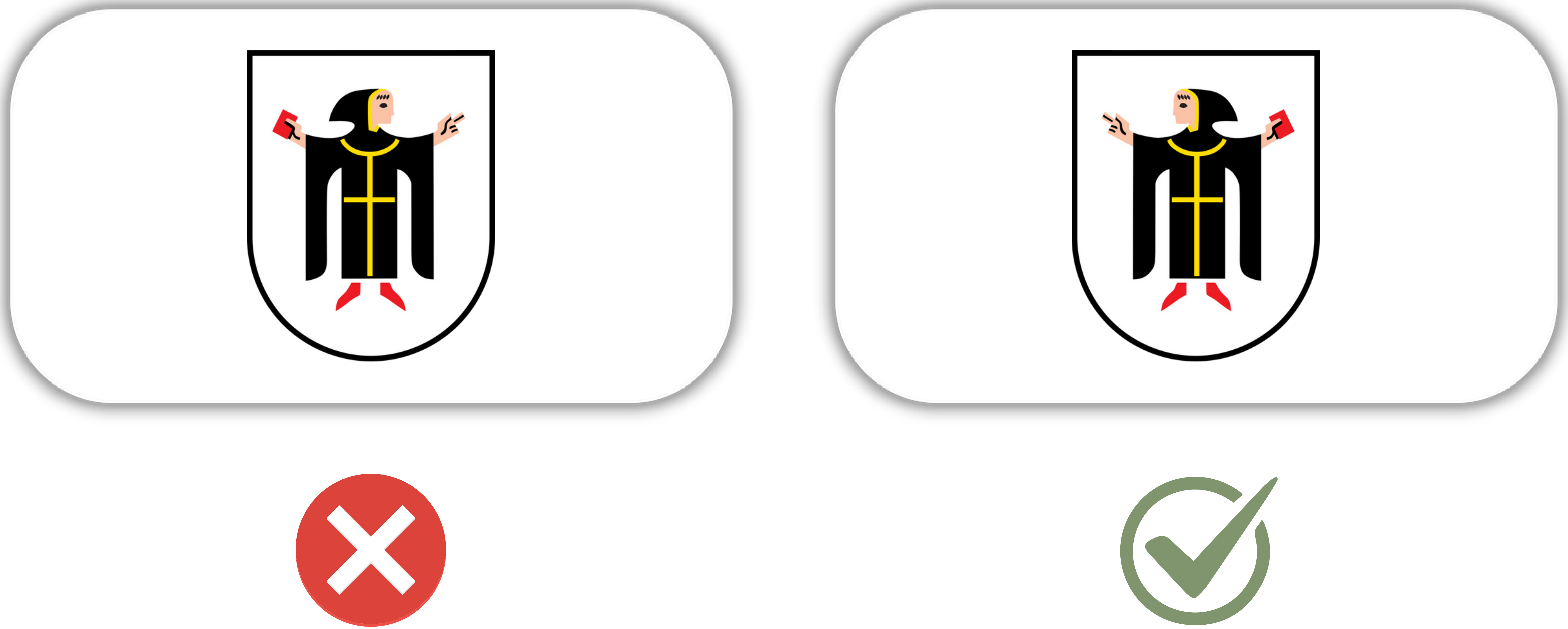
FC BAYERN MÜNCHEN


FC BAYERN MÜNCHEN
FC BAYERN MÜNCHEN







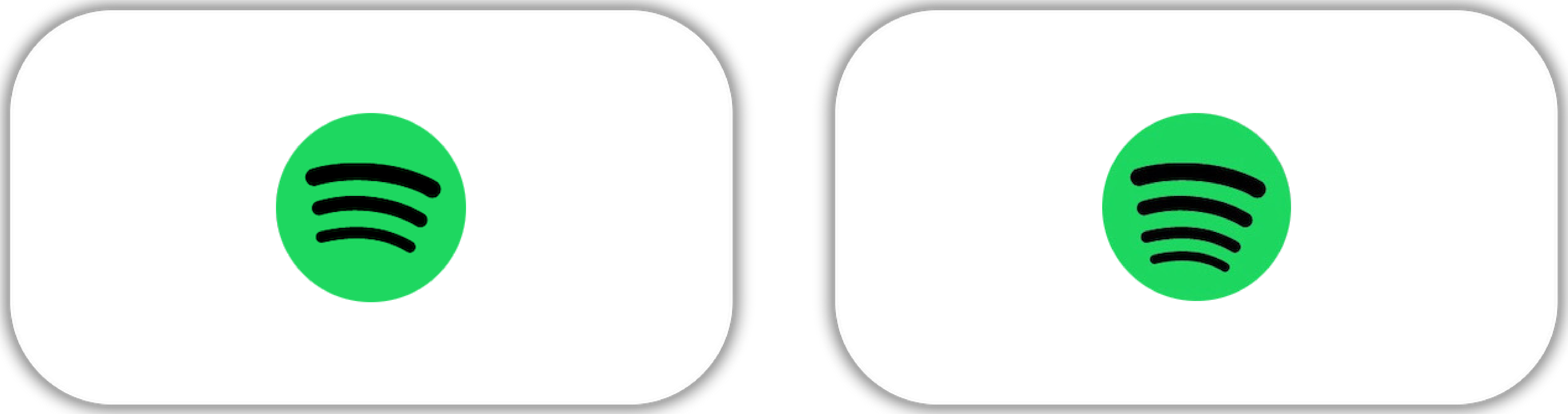


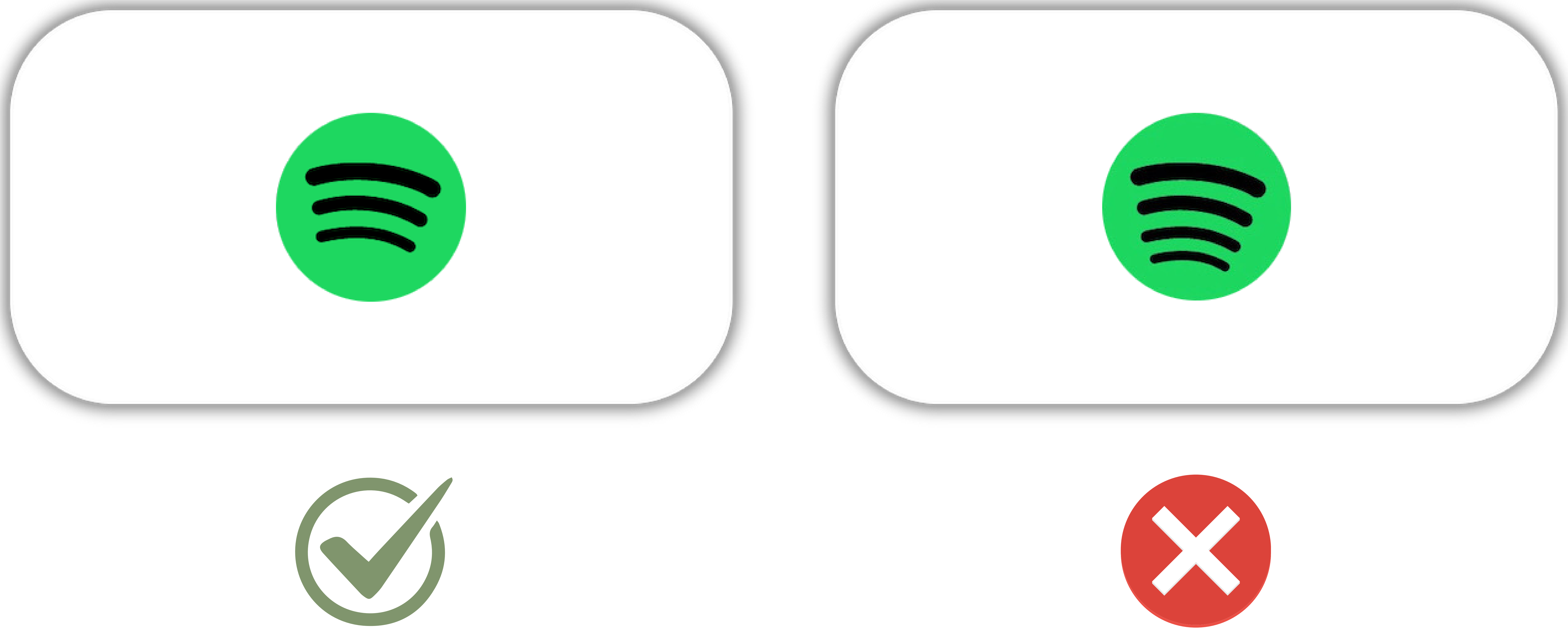
YouTube

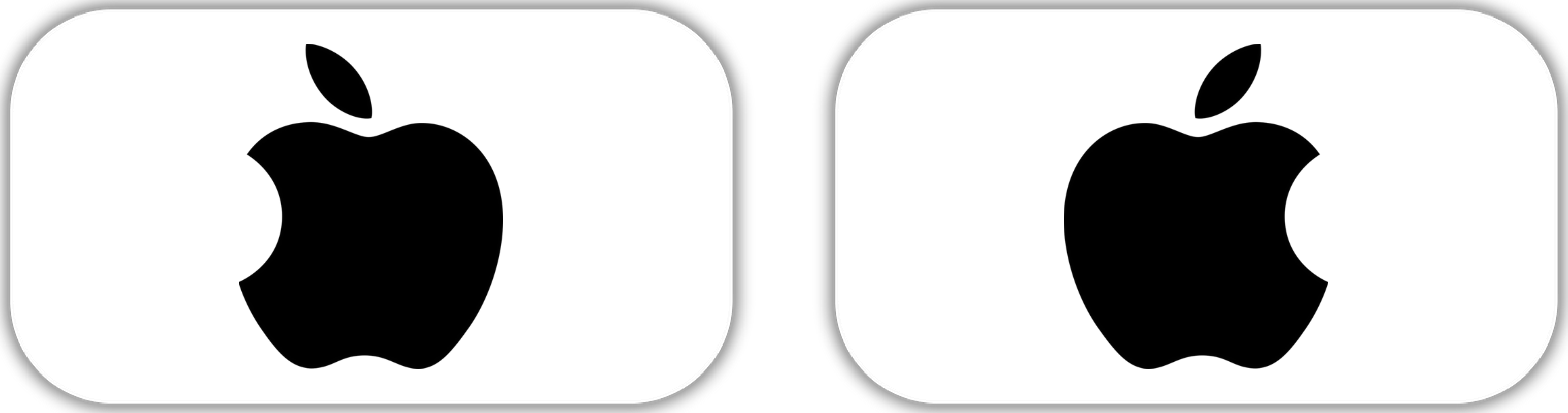
YouTube

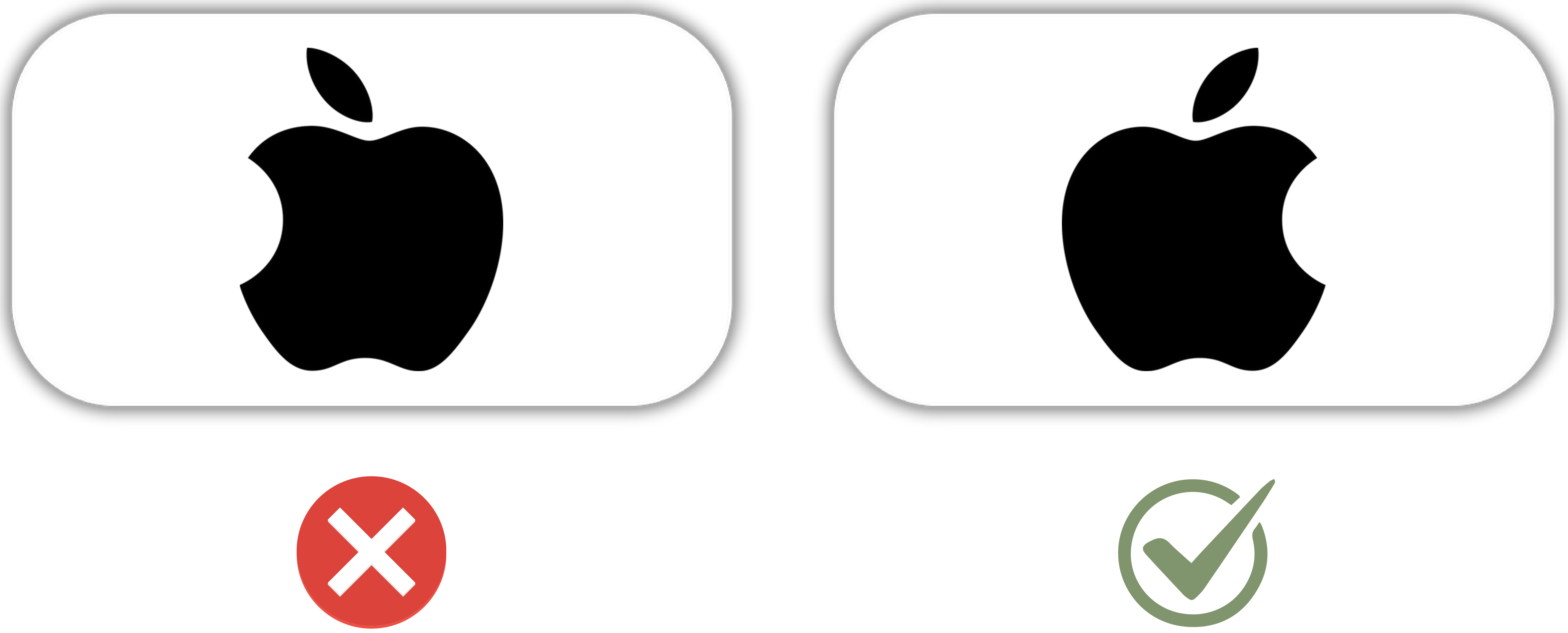
YouTube

YouTube









LIDL

LIDL

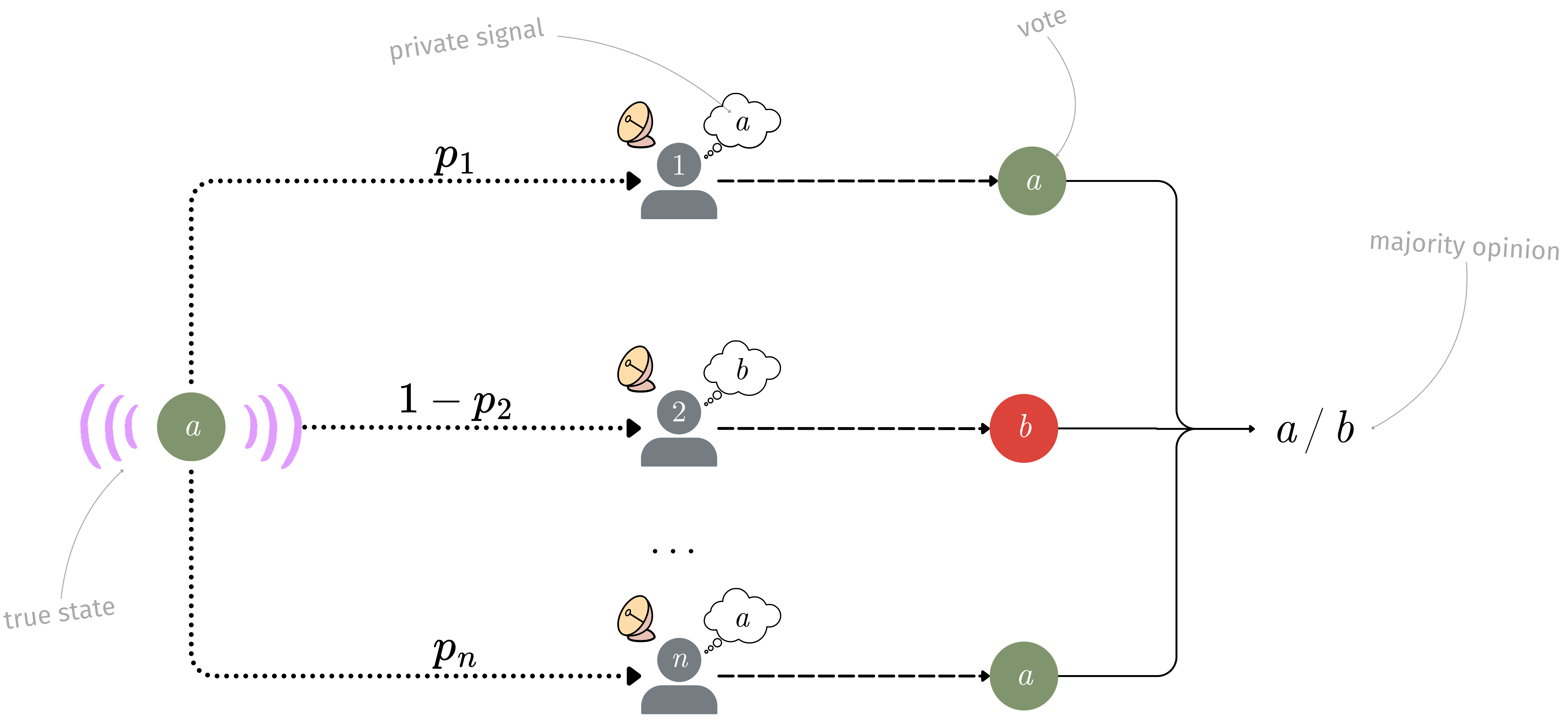

LIDL
LIDL

How did you do?

And how did the group do?

This is the epistemic model we’re thinking in.

AGENTS AS NOISY ESTIMATORS OF THE TRUTH



Note that in the epistemic model it's possible for all voters to vote for the *wrong* alternative.

Unlike the other view of voting, in which the correct alternative is whatever the people want.

We work in a setting where an odd number of *agents* vote on two *alternatives*, one of which is correct.

Each agent has a specific *competence*, which is the probability of voting for the correct alternative.

NOTATION

we assume n is odd so as not to worry about ties

voters	$N = \{1, \ldots, n\}$
alternatives	$A = \{a, b\}$
correct alternative	a
voter i's vote	$v_i \in A$
profile of votes	$\boldsymbol{v} = (v_1, \ldots, v_n)$
voter i's competence	$\Pr[v_i = a] = p_i$, with $p_i \in [0, 1]$
majority vote	$F_{maj}(\boldsymbol{v}) = x$, such that $v_i = x$ for a (strict) majority of voters

i's guess of the right answer

we write profiles as words: (a, a, b, a, ...) $\rightarrow$ $aaba$...

probability that i gets the right answer

CONDORCET

I want to make some assumptions.

ASSUMPTIONS

(Competence) Agents are better than random at being correct:

$$p_i > \frac{1}{2}, \text{ for any voter } i \in N.$$

(Equal Competence) All agents have the same competence:

$$p_i = p_j = p, \text{ for all voters } i, j \in N.$$

(Independence) Voters vote independently of each other:

$$\Pr[v_i = x, v_j = y] = \Pr[v_i = x] \cdot \Pr[v_j = y], \text{ for all voters } i, j \in N.$$

CONDORCET

I claim that under these conditions, the majority tends to get it right!

We want to understand the probability that the majority opinion is correct:

$$\Pr\left[F_{maj}(v_1, \ldots, v_n) = a\right]$$

Computing the probability of a correct majority decision becomes more and more involved as the number of agents grows.

But let's start simple.

ONE VOTER

The profile is $v = (v_1)$.

The probability of a correct decision is:

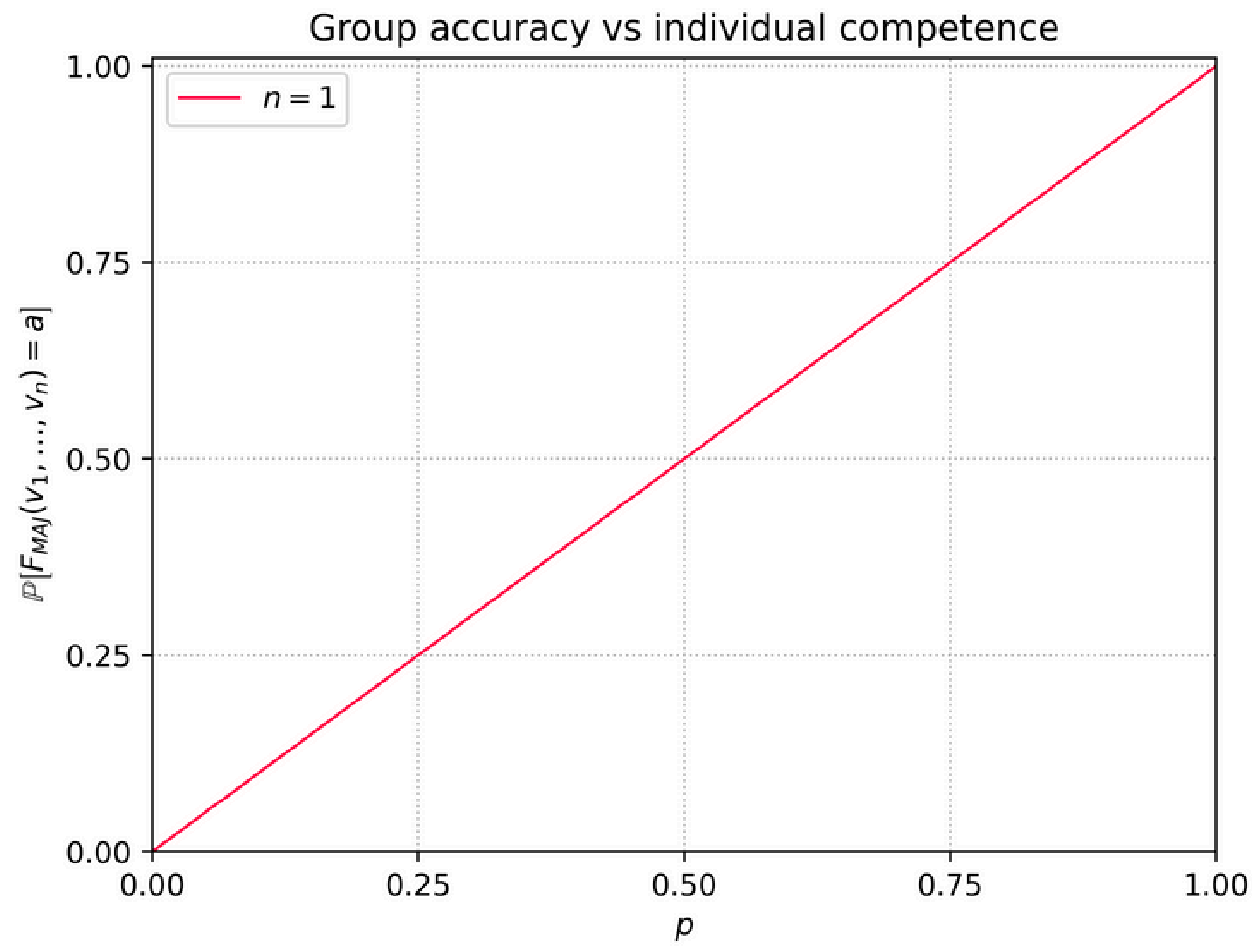
$$\begin{aligned}\Pr\left[F_{maj}(v_1) = a\right] &= \Pr\left[v_1 = a\right] \\ &= p \\ &> 1/2.\end{aligned}$$

by the Equal Competence assumption

by the Competence assumption

Note that as p grows, so does group accuracy.

in this case, trivially



TWO VOTERS

The profile is $v = (v_1, v_2)$.

Oh wait, we're not looking at this case.

THREE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3)$.

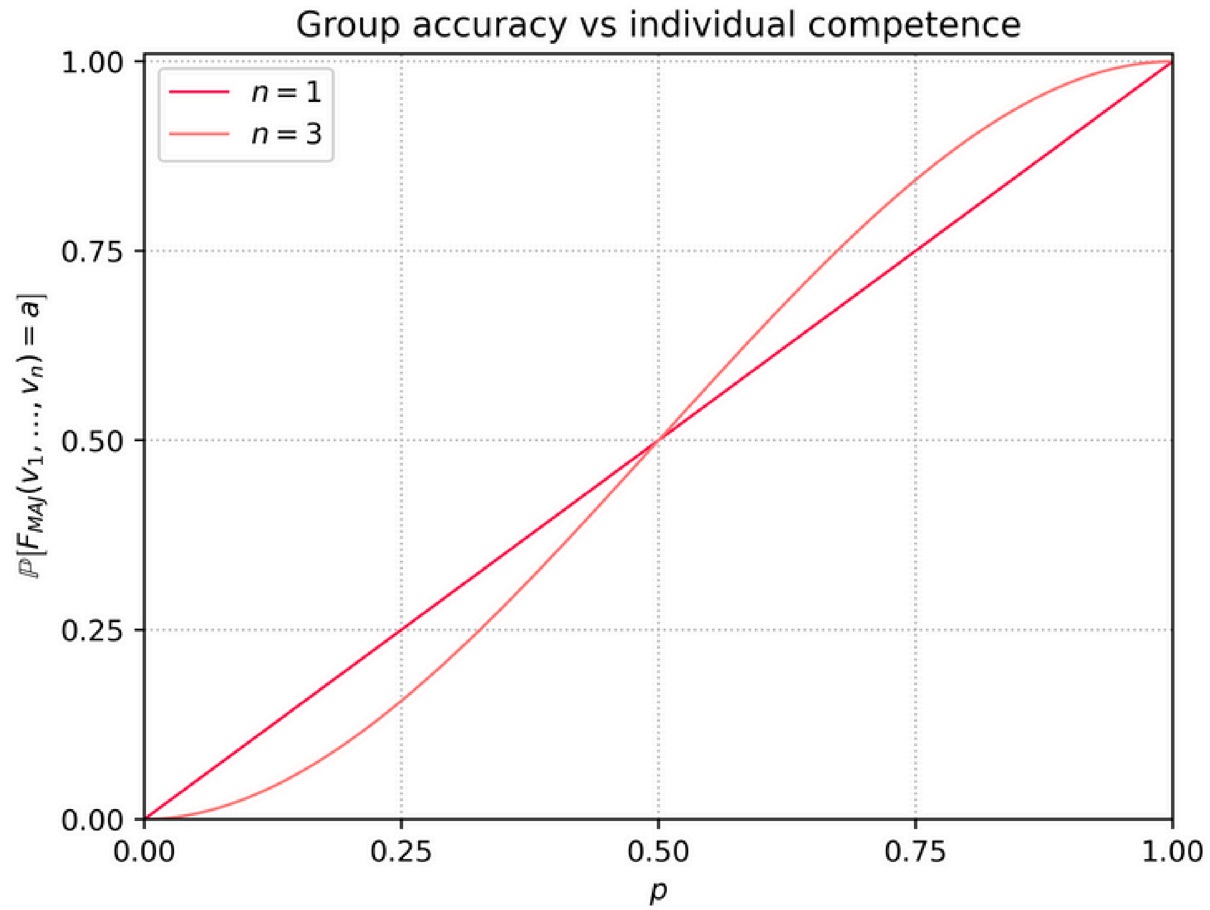
The probability of a correct majority decision is:

$$\begin{aligned}
\Pr\left[F_{maj}(\boldsymbol{v}) = a\right] &= \Pr\left[\text{a majority of voters in } \boldsymbol{v} \text{ are correct}\right] \\
&= \Pr\left[\boldsymbol{v} \text{ is either } aab, aba, baa, \text{ or } aaa\right] \\
&= \Pr\left[\boldsymbol{v} = aab\right] + \Pr\left[\boldsymbol{v} = aba\right] + \Pr\left[\boldsymbol{v} = baa\right] + \Pr\left[\boldsymbol{v} = aaa\right] \\
&= \Pr\left[v_1 = a\right] \cdot \Pr\left[v_2 = a\right] \cdot \Pr\left[v_3 = b\right] + \\
&\quad\ \Pr\left[v_1 = a\right] \cdot \Pr\left[v_2 = b\right] \cdot \Pr\left[v_3 = a\right] + \\
&\quad\ \Pr\left[v_1 = b\right] \cdot \Pr\left[v_2 = a\right] \cdot \Pr\left[v_3 = a\right] + \\
&\quad\ \Pr\left[v_1 = a\right] \cdot \Pr\left[v_2 = a\right] \cdot \Pr\left[v_3 = a\right] \\
&= p \cdot p \cdot (1-p) + p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot p + p \cdot p \cdot p \\
&= 3p^2(1-p) + p^3 \\
&> p.
\end{aligned}$$

by the Independence assumption

by the Equal Competence assumption

by the Competence assumption



Note that as p grows, so does group accuracy.

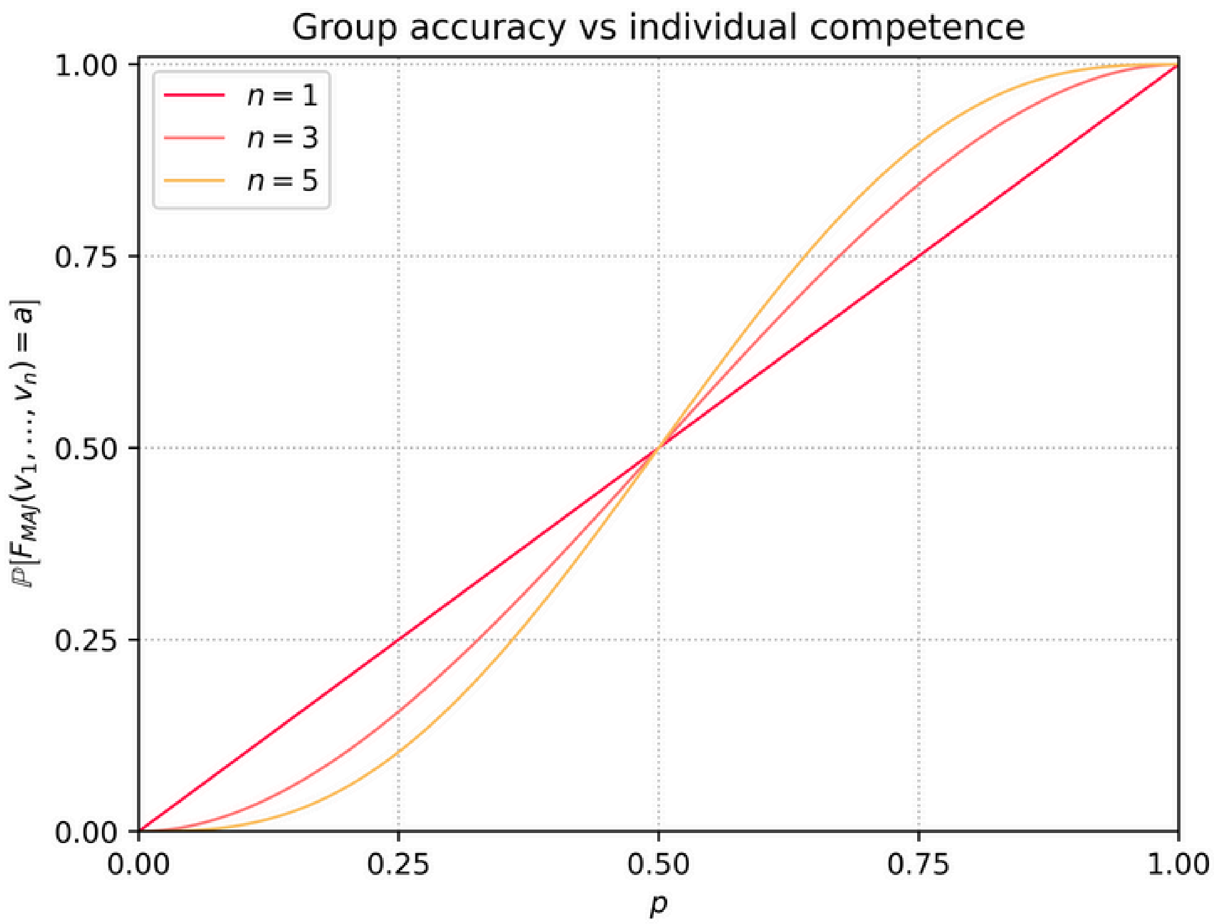
And a group of size 3 is more likely to be correct than a group of size 1.

FIVE VOTERS

The profile is $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5)$.

The probability of a correct majority decision is:

$$\begin{aligned}
\Pr\left[F_{maj}(\boldsymbol{v}) = a\right] &= \Pr\left[\text{a majority of voters in } \boldsymbol{v} \text{ are correct}\right] \\
&= \Pr\left[\text{either exactly } 3, 4 \text{ or } 5 \text{ voters in } \boldsymbol{v} \text{ are correct}\right] \\
&= \Pr\big[\boldsymbol{v} \text{ is either } aaabb, aabab, abaab, baaab, aabba, ababa, baaba \\
&\qquad aaaab, aaaba, aabaa, abaaa, baaaa, \\
&\qquad \text{or } aaaaa\big] \\
&\cdots \\
&= 10 \cdot p^3(1-p)^2 + 5 \cdot p^4(1-p) + p^5 \\
&= \binom{5}{3} \cdot p^3(1-p)^2 + \binom{5}{4} \cdot p^4(1-p)^1 + \binom{5}{5} \cdot p^5
\end{aligned}$$



Note, again, that as p grows, so does group accuracy.

And a group of size 5 is more likely to be correct than a group of size 3.

ANY ODD NUMBER OF VOTERS

The profile is $\boldsymbol{v} = (v_1, \ldots, v_n)$.

The probability of a correct majority decision is:

$$\begin{aligned}
\Pr[F_{maj}(\boldsymbol{v}) = a] &= \Pr[\text{a majority of voters in } \boldsymbol{v} \text{ are correct}] \\
&= \Pr[\text{either exactly } \lfloor n/2 \rfloor + 1, \ldots, n-1, \text{ or } n \text{ voters in } \boldsymbol{v} \text{ are correct}] \\
&= \binom{n}{\lfloor n/2 \rfloor + 1} \cdot p^{\lfloor n/2 \rfloor + 1} (1-p)^{n - (\lfloor n/2 \rfloor + 1)} + + \binom{n}{n-1} p^{n-1} (1-p)^1 + \binom{n}{n} p^n \\
&= \sum_{i = \lfloor n/2 \rfloor + 1}^{n} \binom{n}{i} p^i (1-p)^{n-i}.
\end{aligned}$$

CONDORCET

By the croissants of my ancestors: I claim that the larger the group, the more accurate it is!

And that in the limit, groups are infallible.

Provided there are no dumdums and people make their minds up independently.

THE CONDORCET JURY THEOREM (CJT)

THEOREM

For n voters, with n odd, all of whom have accuracy $p > 1/2$ and vote independently of each other, it holds that:

(1) The accuracy of the group improves as the size of the group grows:

$$\Pr\left[F_{maj}(v_1, \ldots, v_{n+2}) = a\right] > \Pr\left[F_{maj}(v_1, \ldots, v_n) = a\right].$$

(2) The accuracy of the group is at least as good as the accuracy of the inidividual members:

$$\Pr\left[F_{maj}(v_1, \ldots, v_n) = a\right] \geq p.$$

(3) As n goes to infinity, the accuracy of the group approaches 1 asymptotically:

$$\lim_{n \to \infty} \Pr\left[F_{maj}(v_1, \ldots, v_n) = a\right] = 1.$$

How do we prove this?

For one, it's easier to keep track of things using random variables.

indicator random variable	$X_i = \begin{cases} 1, & \text{if voter } i \text{ is correct, i.e., if } v_i = a, \\ 0, & \text{otherwise.} \end{cases}$
sum random variable	$S_n = X_1 + \cdots + X_n$

The events we are interested in are easy to write using random variables.

The probability of a correct/incorrect decision is:

$$\Pr[X_i = 1] = p, \qquad \Pr[X_i = 0] = 1 - p.$$

$S_n = k$ means exactly k voters are correct.

$S_n > \lfloor n/2 \rfloor$ means the majority is correct.

We're aiming to show that $\lim_{n \to \infty} \Pr[S_n > \lfloor n/2 \rfloor] = 1$.

$$\boxed{a \; a} \; b \; b \; a$$

$$X_1 + X_2 + X_5 = 3$$

JACOB BERNOULLI

The variables, may I humbly point out, are called
Bernoulli variables.

To prove that larger groups get better, we derive a recurrence relation for the accuracy of a group of n voters in terms of the accuracy of a group of $n - 2$ voters.

Take $n = 5$.

FIVE VOTERS & A CORRECT MAJORITY

Separate the first two voters, and let's count the ways of obtaining a correct majority.

$\boxed{\_\ \_}\ \_\ \_\ \_$

If the first two voters are wrong, the remaining three have to be correct.

$\boxed{b\ b}\ a\ a\ a$

If exactly one of the first two voters is correct (which can happen in two ways), at least two of the remaining voters have to be correct.

$\boxed{a\ b}\ a\ a\ a$ $\quad \boxed{a\ b}\ a\ a\ b$ $\quad \boxed{a\ b}\ a\ b\ a$ $\quad \boxed{a\ b}\ b\ a\ a$

$\boxed{b\ a}\ a\ a\ a$ $\quad \boxed{b\ a}\ a\ a\ b$ $\quad \boxed{b\ a}\ a\ b\ a$ $\quad \boxed{b\ a}\ b\ a\ a$

If both of the first two voters are correct, at least one of the remaining voters has to be correct.

$\boxed{a\ a}\ a\ a\ a$

$\boxed{a\ a}\ a\ a\ b$ $\quad \boxed{a\ a}\ a\ b\ a$ $\quad \boxed{a\ a}\ b\ a\ a$

$\boxed{a\ a}\ a\ b\ b$ $\quad \boxed{a\ a}\ b\ a\ b$ $\quad \boxed{a\ a}\ b\ b\ a$

The probability of a correct majority is thus:

$$\Pr[S_5 > 2] = (1-p)^2 \cdot \Pr[S_3 > 2] + 2p(1-p) \cdot \Pr[S_3 > 1] + p^2 \cdot \Pr[S_3 > 0]$$

The general version of this recurrence looks as follows.

PROOF OF CLAIM 1: LARGER GROUPS HAVE BETTER ACCURACY

Generalizing the previous identity we get the following recurrence:

$$\Pr\left[S_{n+2} > \left\lfloor \frac{n+2}{2} \right\rfloor\right] = (1-p)^2 \cdot \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor + 1\right] + 2p(1-p) \cdot \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor\right] + p^2 \cdot \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor - 1\right] \quad (1)$$

The events on the right-hand-side can be rewritten as:

$$\Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor - 1\right] = \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor\right] + \binom{n}{\lfloor n/2 \rfloor} \cdot p^{\lfloor n/2 \rfloor}(1-p)^{\lfloor n/2 \rfloor + 1} \quad (2)$$

$$\Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor + 1\right] = \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor\right] - \binom{n}{\lfloor n/2 + 1 \rfloor} \cdot p^{\lfloor n/2 \rfloor + 1}(1-p)^{\lfloor n/2 \rfloor} \quad (3)$$

Plug Equalities 2 and 3 in 1, and write $\binom{n}{\lfloor n/2 \rfloor} = \binom{n}{\lfloor n/2 \rfloor + 1} = c$:

$$\Pr\left[S_{n+2} > \left\lfloor \frac{n+2}{2} \right\rfloor\right] = \Pr\left[S_n > \left\lfloor \frac{n}{2} \right\rfloor\right] + c \cdot p^{\lfloor n/2 \rfloor + 1}(1-p)^{\lfloor n/2 \rfloor + 1}(2p-1).$$

Since $1/2 < p < 1$, the second term on the right-hand side is positive.

PROOF OF CLAIM 2: THE GROUP IS BETTER THAN ITS MEMBERS

This follows from Claim 1:

$$
\begin{aligned}
p &= \Pr[S_1 > 0] \\
&< \Pr[S_3 > 1] \\
&\cdots \\
&< \Pr\left[S_n > \lfloor {}^{n}\!/\!{}_{2} \rfloor\right] \\
&\cdots
\end{aligned}
$$

Claim 3, i.e., that in the limit accuracy is 1, follows from *The Law of Large Numbers*.

JACOB BERNOULLI

The intuition for the law of large numbers is as follows.

Say we have random variables that take value 1 with probability 0.02, and 0 with probability 0.98.

The expected value of such a variable is 0.02.

Now, if we sample a million such variables independently, then we'd expect around 2% (i.e., 20000) of them to have value 1.

More to the point, we'd expect the average over many samples to be around 0.02.

The sample mean approaches the expected value, i.e., the 'true', theoretical mean.

$$X_i = \begin{cases} 1, & \text{with probability } 0.02, \\ 0, & \text{with probability } 0.98, \end{cases}$$

$$\mathbb{E}[X_i] = 1 \cdot 0.02 + 0 \cdot 0.98 = 0.02.$$

THE OX

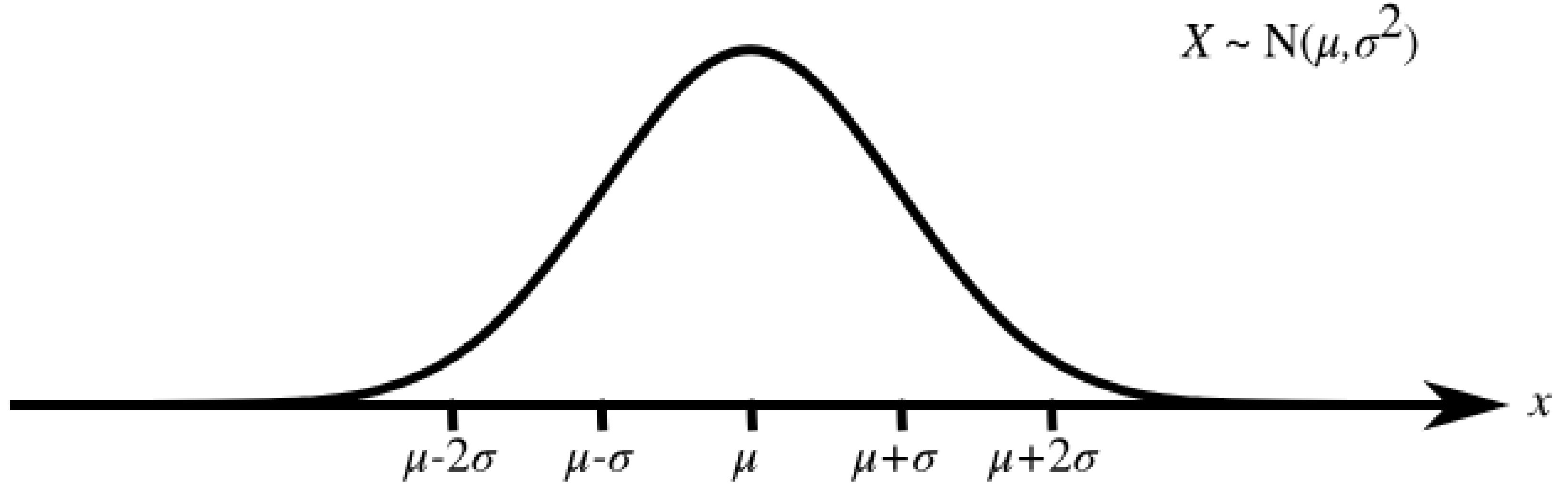
This probably explains what happened at the Plymouth county fair!

Suppose farmers' guesses are distributed like this:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Suppose farmers' guesses are distributed like this:

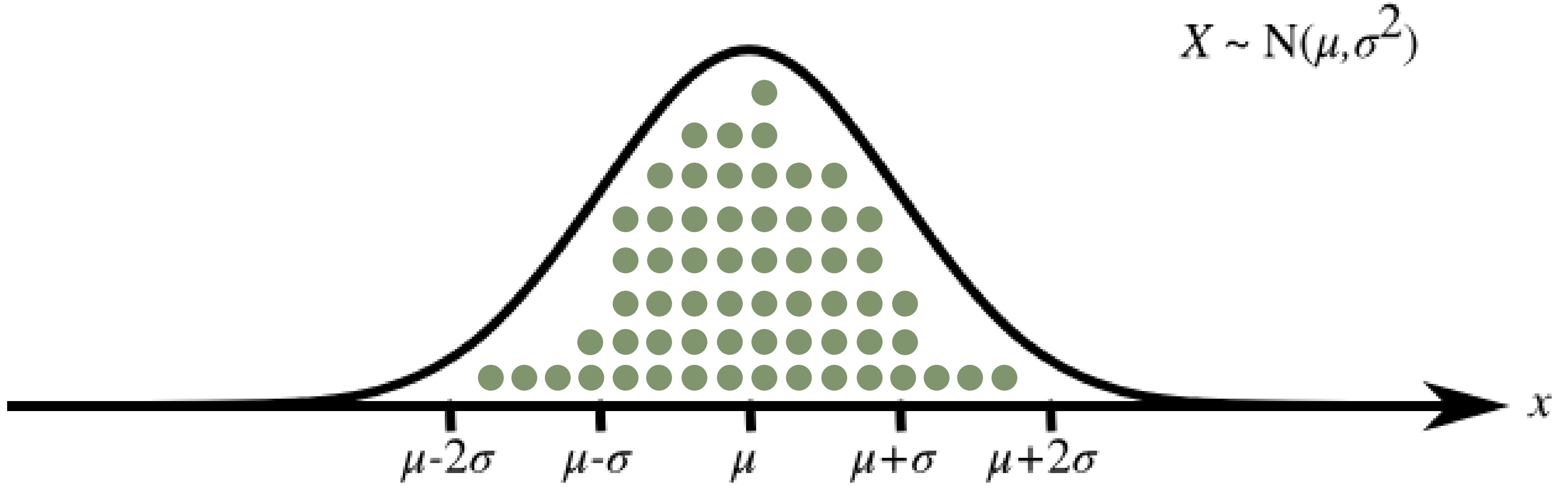
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Suppose farmers' guesses are distributed like this:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Then we're likely to see the approximately 800-size sample approximate this.



With the Condorcet Jury Theorem, we expect a small bias for the truth to lead to more votes for the correct alternative.

THE WEAK LAW OF LARGE NUMBERS

THEOREM

If $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables such that $\mathbb{E}[X_i] = \mu$, then, for any $\varepsilon > 0$, it holds that:

$$\lim_{n \to \infty} \Pr\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right] = 1.$$

PROOF OF CLAIM 3: IN THE LIMIT, ACCURACY IS PERFECT

We need to show that:

$$\lim_{n\to\infty} \Pr\left[S_n > \lfloor n/2 \rfloor\right] = 1.$$

Now, the expected value (i.e., mean μ) of the voter random variables X_i is:

$$\mathbb{E}[X_i] = 1 \cdot p + 0 \cdot (1-p) = p,$$

and the Weak Law of Large Numbers gives us that, for any $\varepsilon > 0$:

$$\lim_{n\to\infty} \Pr\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] = 1.$$

Choosing ε appropriately and massaging this expression we obtain the desired conclusion.

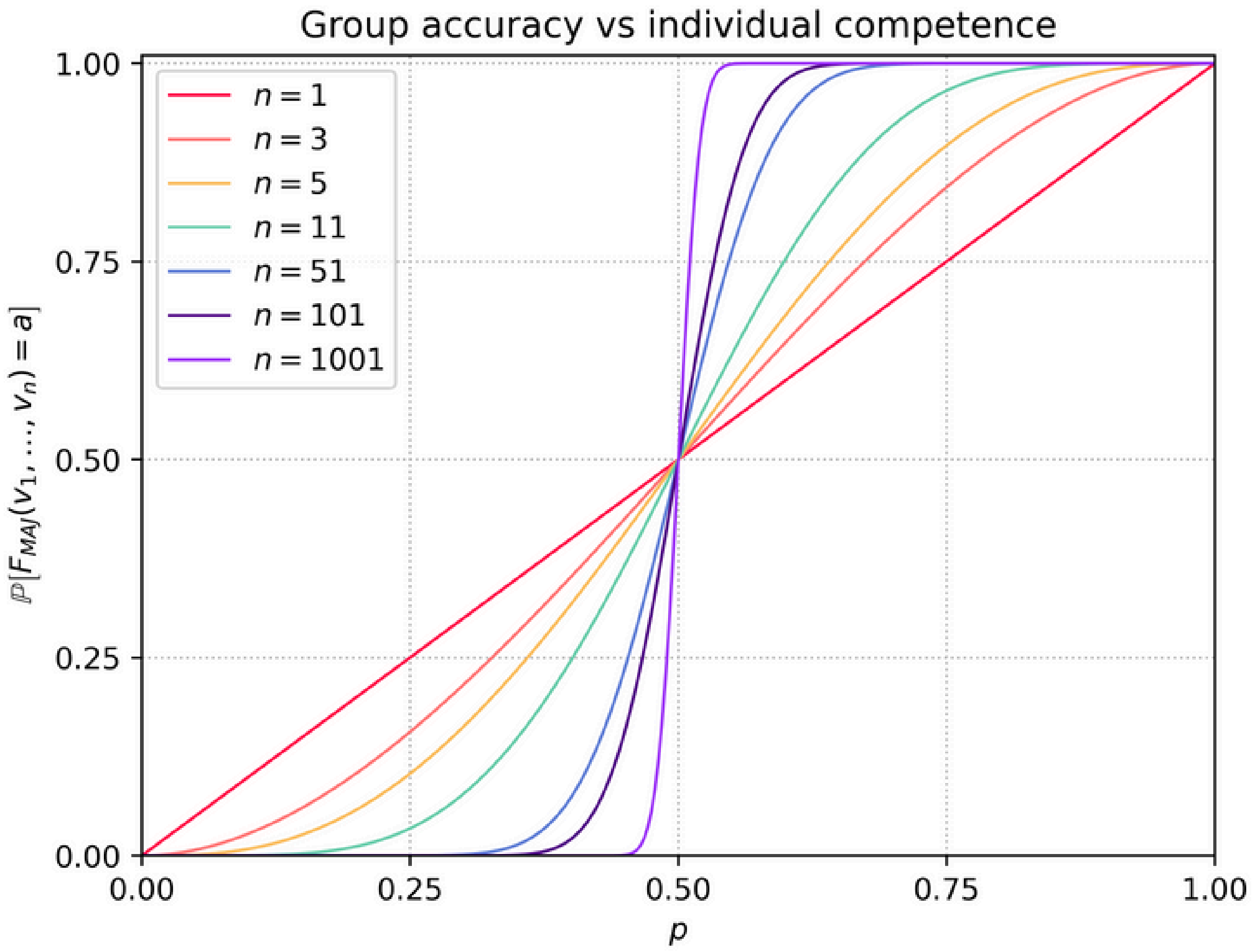
Let's sum up.

CONDORCET

Groups are better than their members.

The larger the group, the better.

In the limit, performance is perfect.



CONDORCET

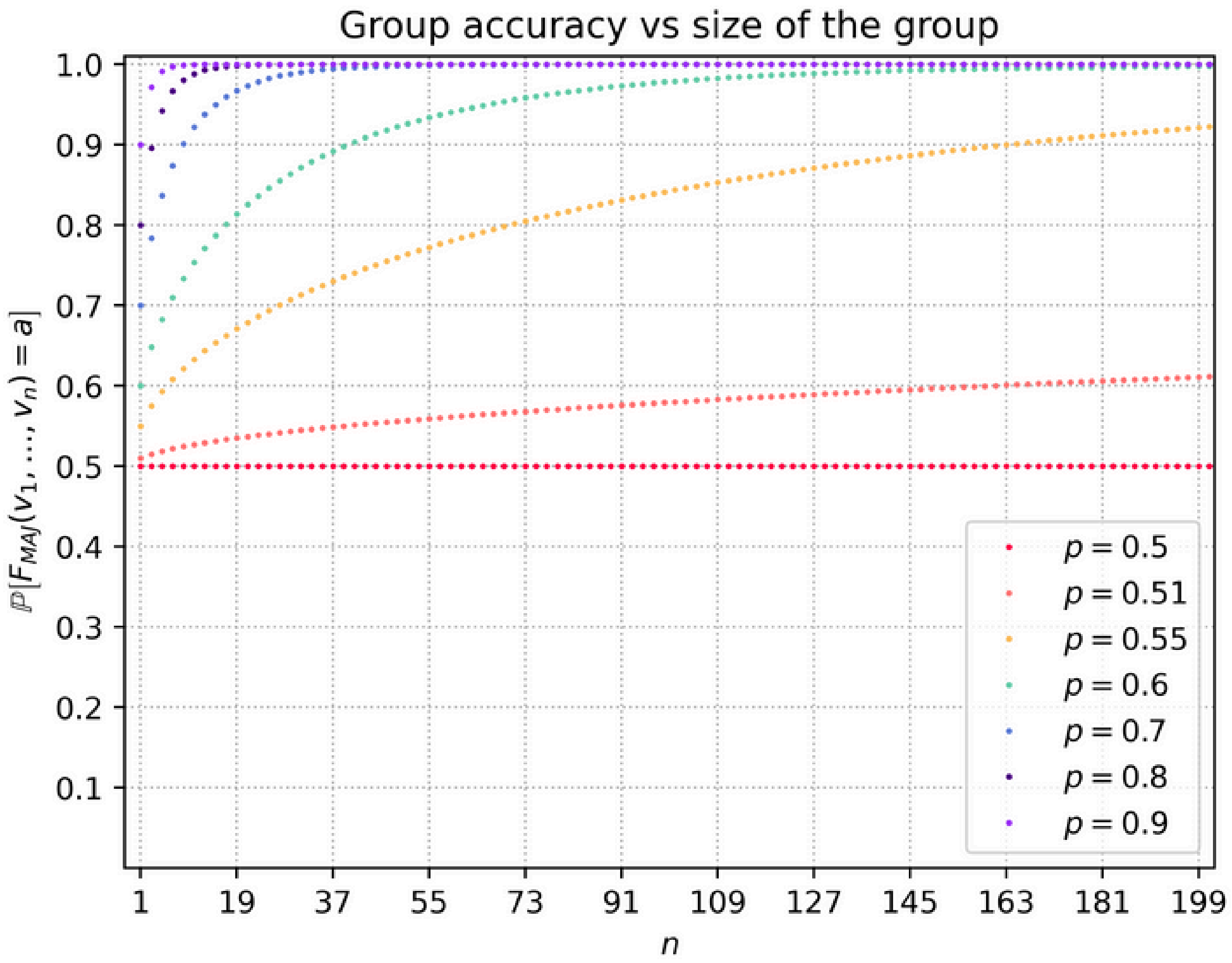
Groups are better than their members.

The larger the group, the better.

In the limit, performance is perfect.

And performance grows fast with the size of the group.

Provided $p > 0.5$.



CONDORCET

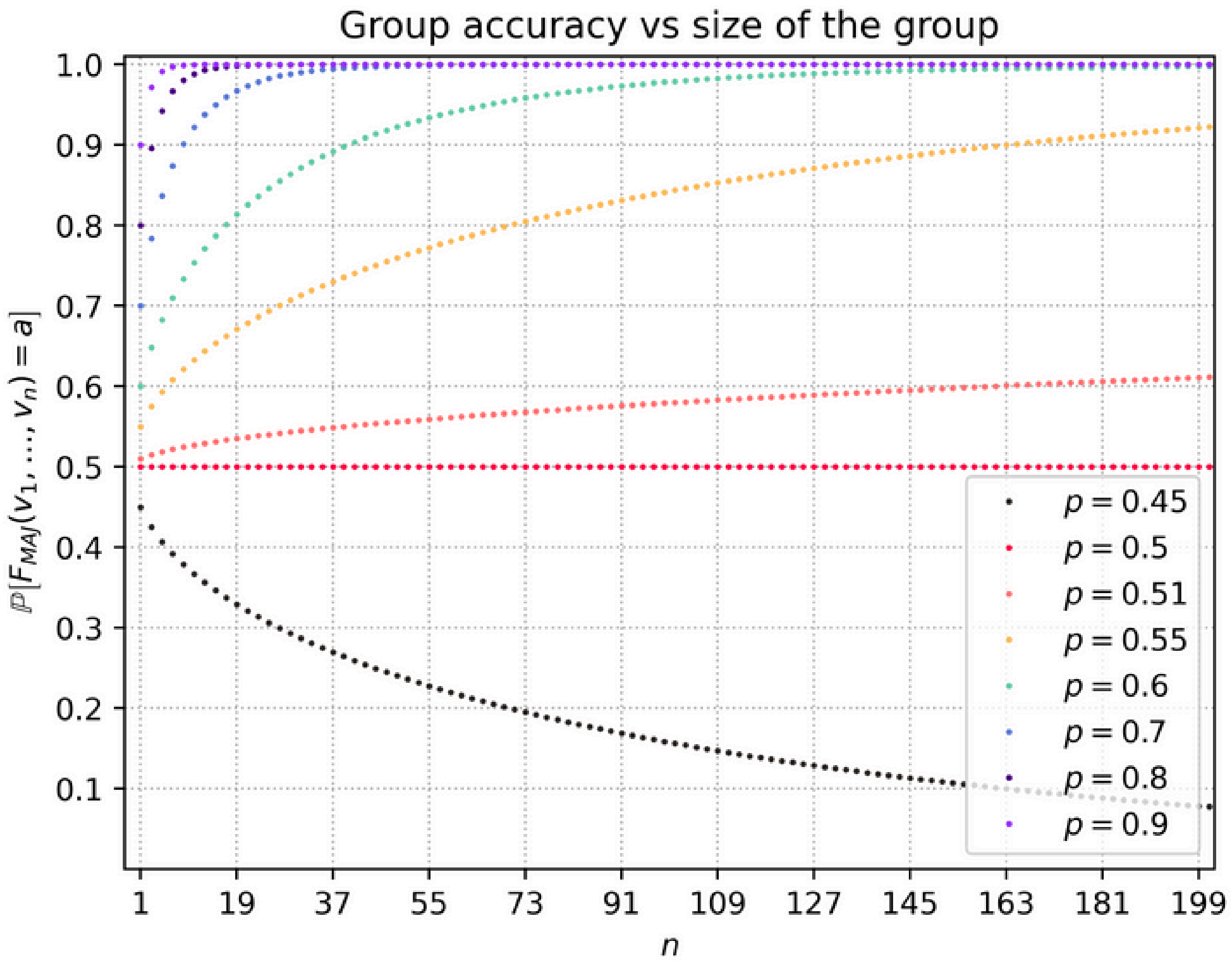
Groups are better than their members.

The larger the group, the better.

In the limit, performance is perfect.

And performance grows fast with the size of the group.

Provided p > 0.5.



Interestingly, actual juries don’t operate at all according to the conditions of the Condorcet Jury Theorem.

Interestingly, actual juries don’t operate at all according to the conditions of the Condorcet Jury Theorem.

How *do* actual juries work, by the way?

🔎