Individual preferences during classification

align with

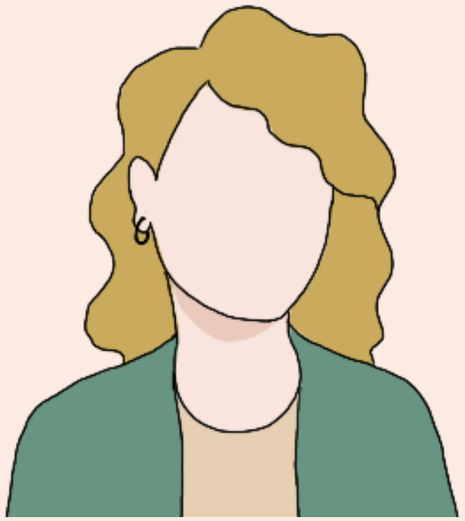
kin term semantics across languages

Maisy Hallam, Simon Kirby,
Fiona M. Jordan and Kenny Smith

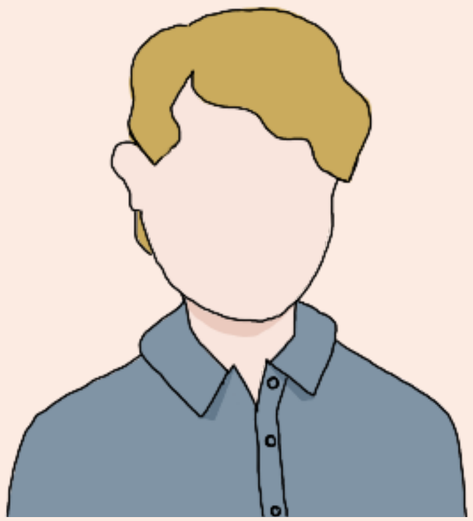
A metric for quantifying
semantic encoding in a category system,

cross-linguistic constraints on the
semantics of kinship terminology,

and experimental evidence for a **cognitive bias** that reflects cross-linguistic constraints.



“sister”



“brother”

English speakers distinguish their siblings by gender.

“kakak”
(older sibling)

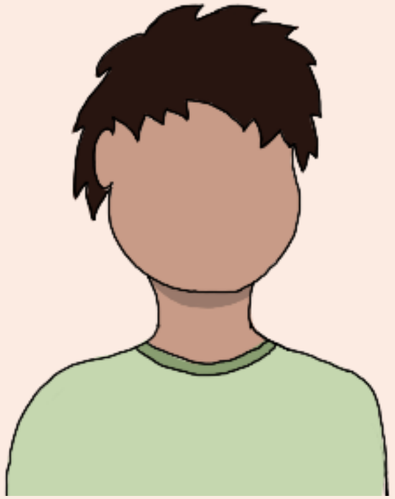
“adik”
(younger sibling)

But **Indonesian** speakers distinguish their siblings by seniority.

“dīdī”
(older sister)

“behna”
(younger sister)

“bhaiya”
(older brother)



“choṭā bhāī”
(younger brother)

And **Hindi** speakers distinguish by gender **and** seniority!

Across the full family tree, the scope of semantic diversity increases exponentially.

Maisy Hallam ✧ Centre for Language Evolution

Across the full family tree, the scope of semantic diversity increases exponentially.

Gender

a relative’s gender

Seniority

is this relative older
or younger?

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Gender of Connecting Relative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

There are (at least!) six semantic features that distinguish kin cross-linguistically.

Maisy Hallam ✧ Centre for Language Evolution

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Gender of Connecting Relative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

There are (at least!) six semantic features that distinguish kin cross-linguistically.

Maisy Hallam ✧ Centre for Language Evolution

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Gender of Connecting Relative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

There are (at least!) six semantic features that distinguish kin cross-linguistically.

Maisy Hallam ✧ Centre for Language Evolution

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Gender of Connecting Relative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

There are (at least!) six semantic features that distinguish kin cross-linguistically.

Maisy Hallam ✧ Centre for Language Evolution

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Gender of **C**onnecting **R**elative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

There are (at least!) six semantic features that distinguish kin cross-linguistically.

Maisy Hallam ✧ Centre for Language Evolution

But languages don’t treat these features equally: we want to know **how much weight** each feature receives in a particular language.

How can we characterise constraints on kin term semantics across the world's languages?

One simple principle:

shared form corresponds to

shared meaning.

(woman's older female sibling)

(man's younger female sibling)

"Shared form corresponds to shared meaning".

(woman’s older female sibling)

“sister”

“sister”

(man’s younger female sibling)

“Shared form corresponds to shared meaning”.

(woman's older female sibling)

(man's younger female sibling)

"sister"

"sister"

Generation

Gender

Lineality

Gender of **C**onnecting **R**elative

Seniority

Speaker Gender

"Shared form corresponds to shared meaning".

(woman’s older female sibling)

(man’s younger female sibling)

“sister”

“sister”

Generation

Gender

Lineality

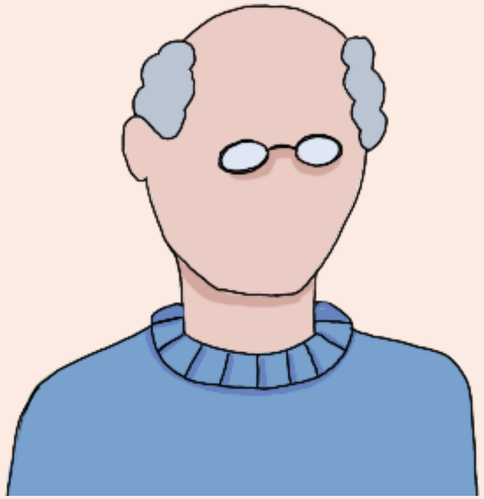
X

X

X

“Shared form corresponds to shared meaning”.

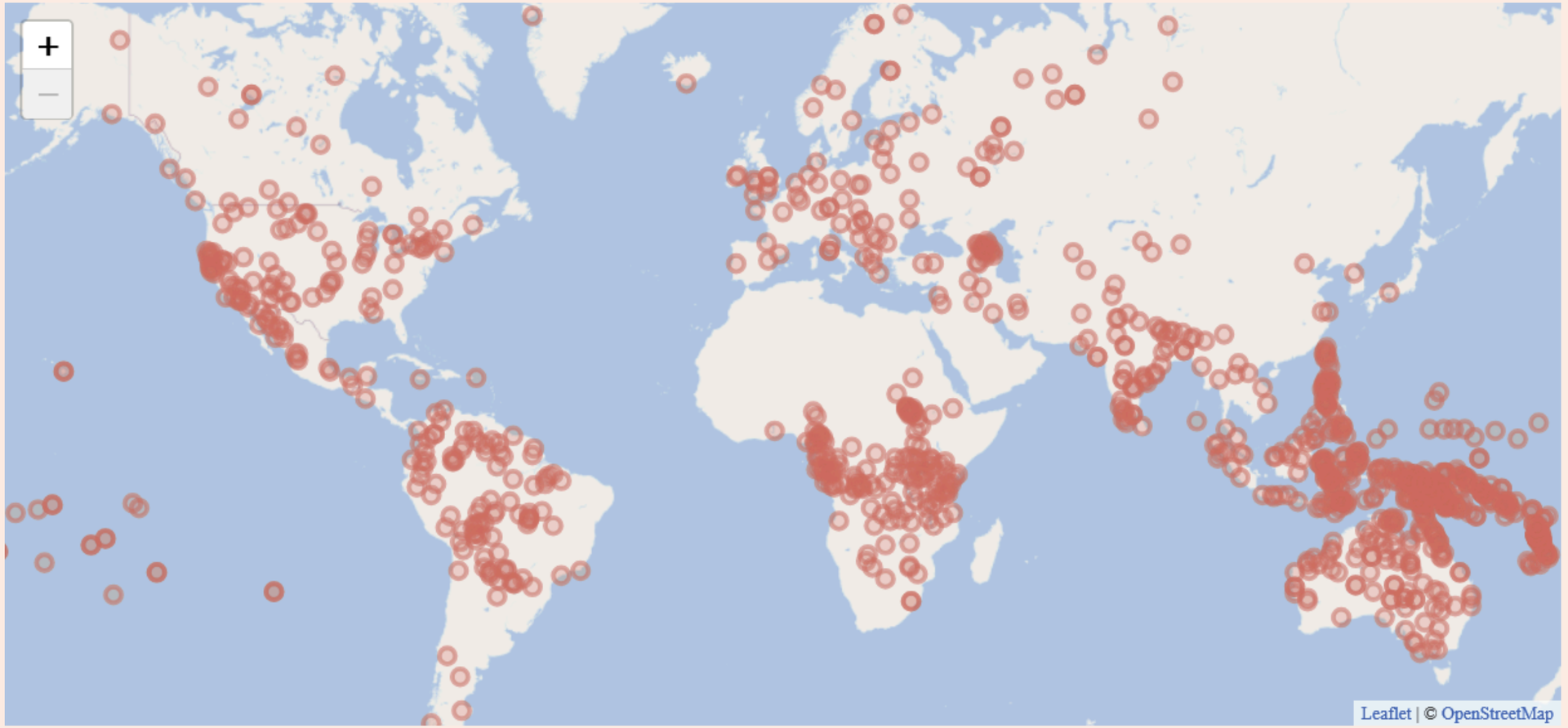
(woman’s mother’s mother)



man’s father’s father)

“**grand**mo**ther**” “**grand**fa**ther**”

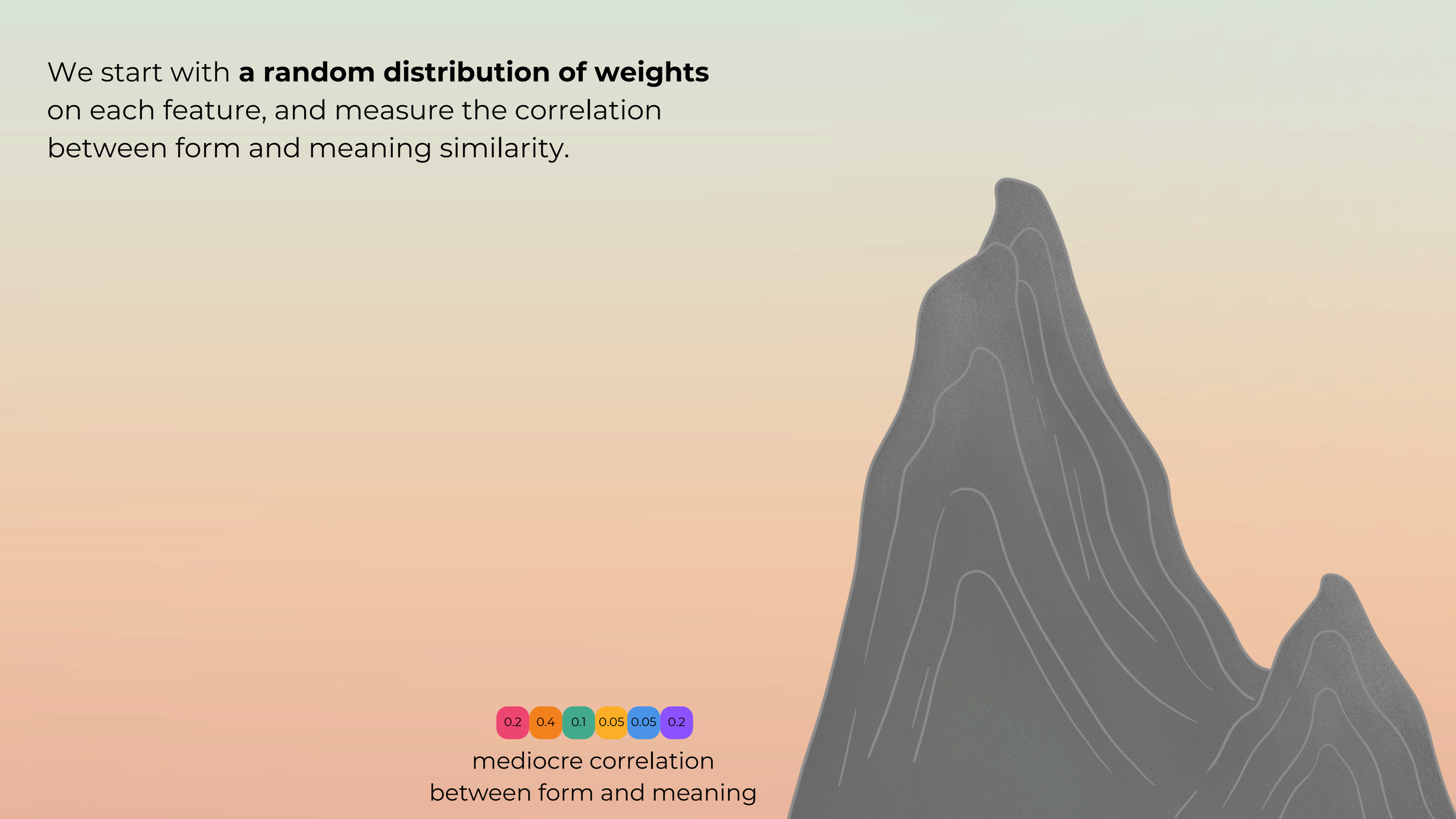
“Shared form corresponds to shared meaning”.



We measured **how meaning and form correspond** in kinship term data from 1200+ global languages

(Kinbank, Passmore et al. 2025)

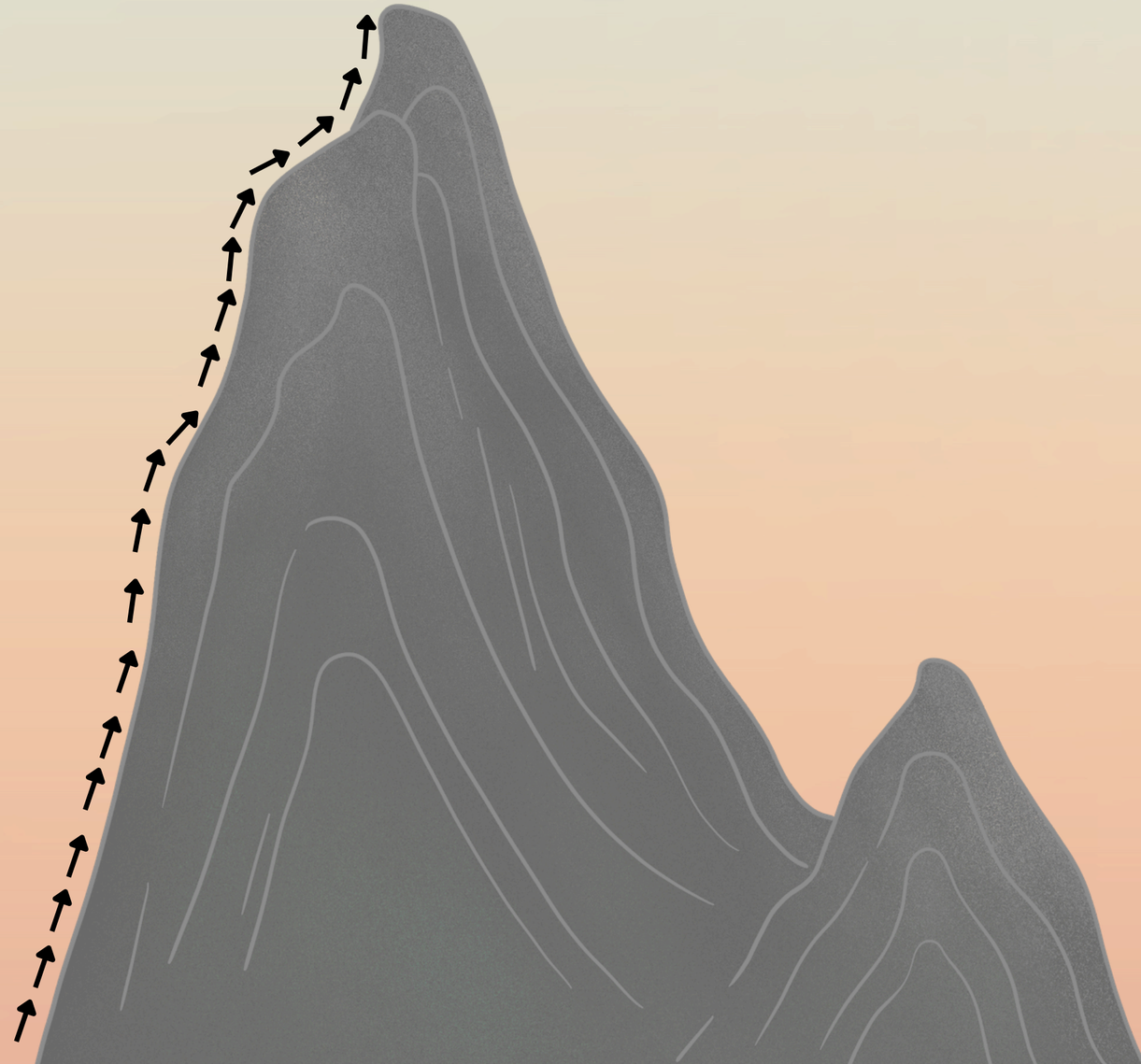
We start with **a random distribution of weights** on each feature, and measure the correlation between form and meaning similarity.



mediocre correlation
between form and meaning

We start with **a random distribution of weights** on each feature, and measure the correlation between form and meaning similarity.

We make **incremental changes** to these weights,



0.2 0.4 0.1 0.05 0.05 0.2

kind of rubbish correlation between form and meaning

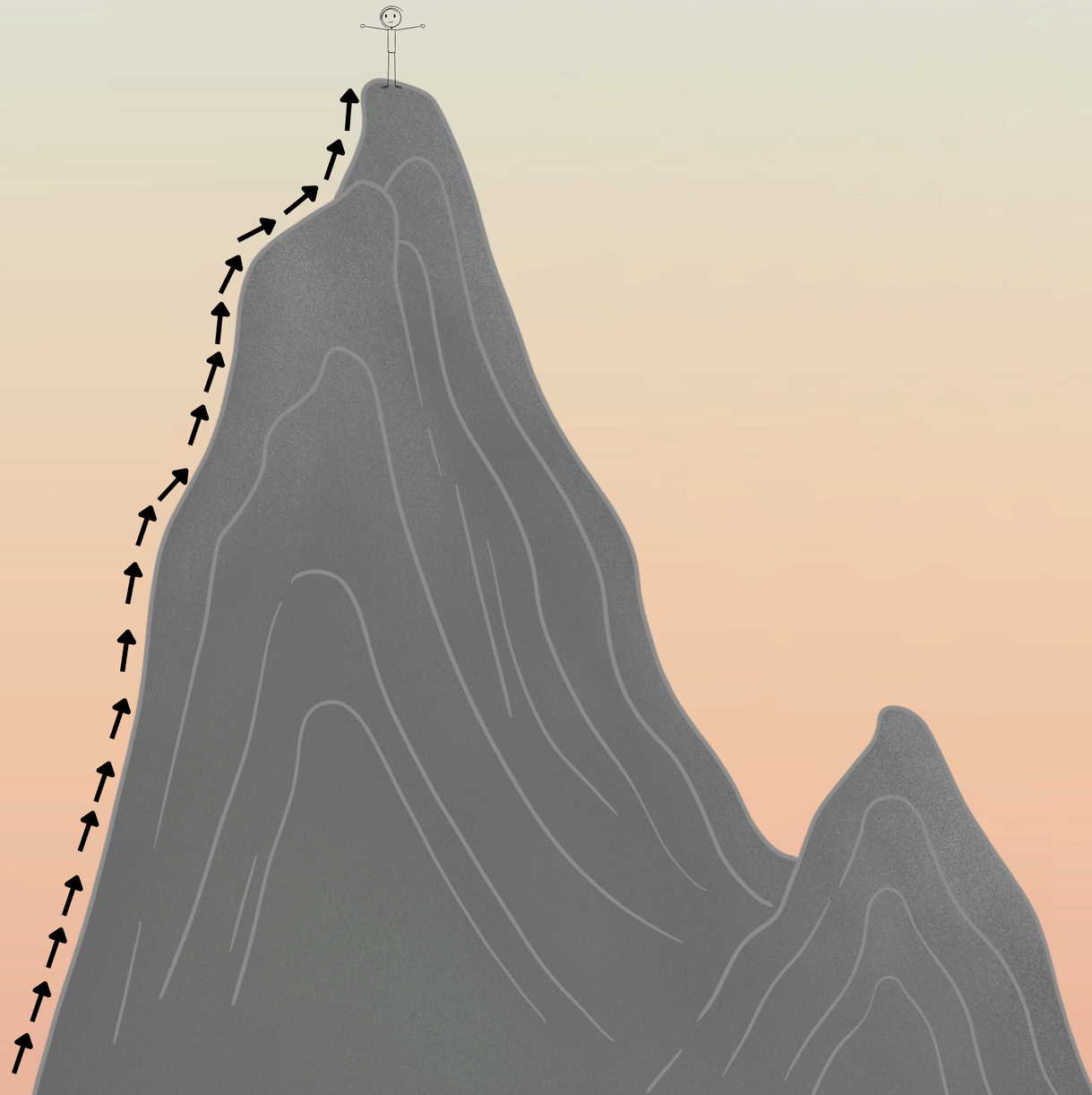
We start with **a random distribution of weights** on each feature, and measure the correlation between form and meaning similarity.

We make **incremental changes** to these weights,

until we achieve the **best possible correlation** between forms and meanings,

0.6 0.2 0.1 0.1 0 0

best correlation between form and meaning!



0.2 0.4 0.1 0.05 0.05 0.2

kind of rubbish correlation between form and meaning

We start with **a random distribution of weights** on each feature, and measure the correlation between form and meaning similarity.

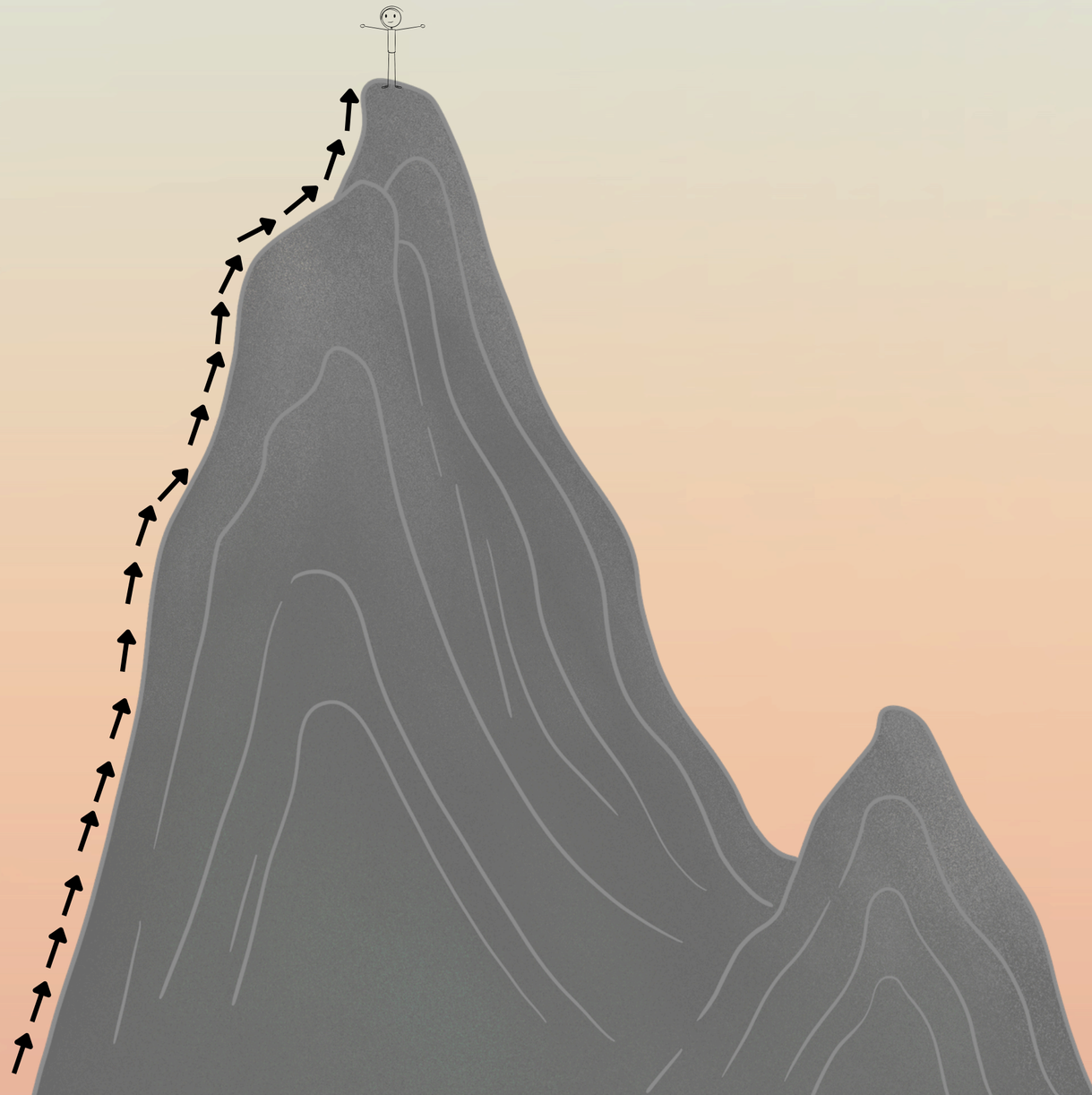
We make **incremental changes** to these weights,

until we achieve the **best possible correlation** between forms and meanings,

giving us an **optimal value** for each feature,

0.6 | 0.2 | 0.1 | 0.1 | 0 | 0

best correlation between form and meaning!



0.2 | 0.4 | 0.1 | 0.05 | 0.05 | 0.2

kind of rubbish correlation between form and meaning

We start with **a random distribution of weights** on each feature, and measure the correlation between form and meaning similarity.

We make **incremental changes** to these weights,

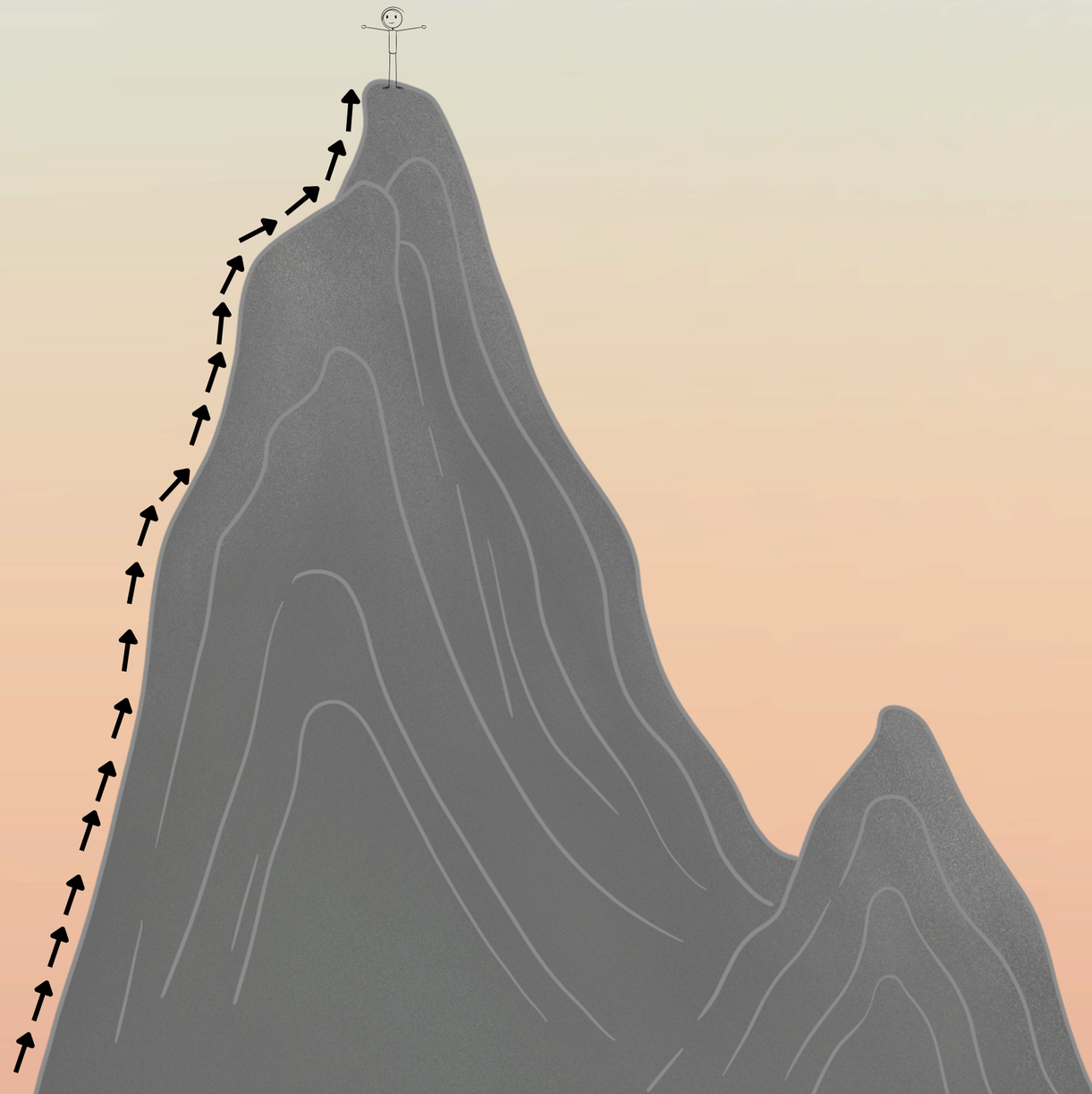
until we achieve the **best possible correlation** between forms and meanings,

giving us an **optimal value** for each feature,

that tells us **how much it contributes** to form-meaning mappings in that language.

0.6 | 0.2 | 0.1 | 0.1 | 0 | 0

best correlation between form and meaning!

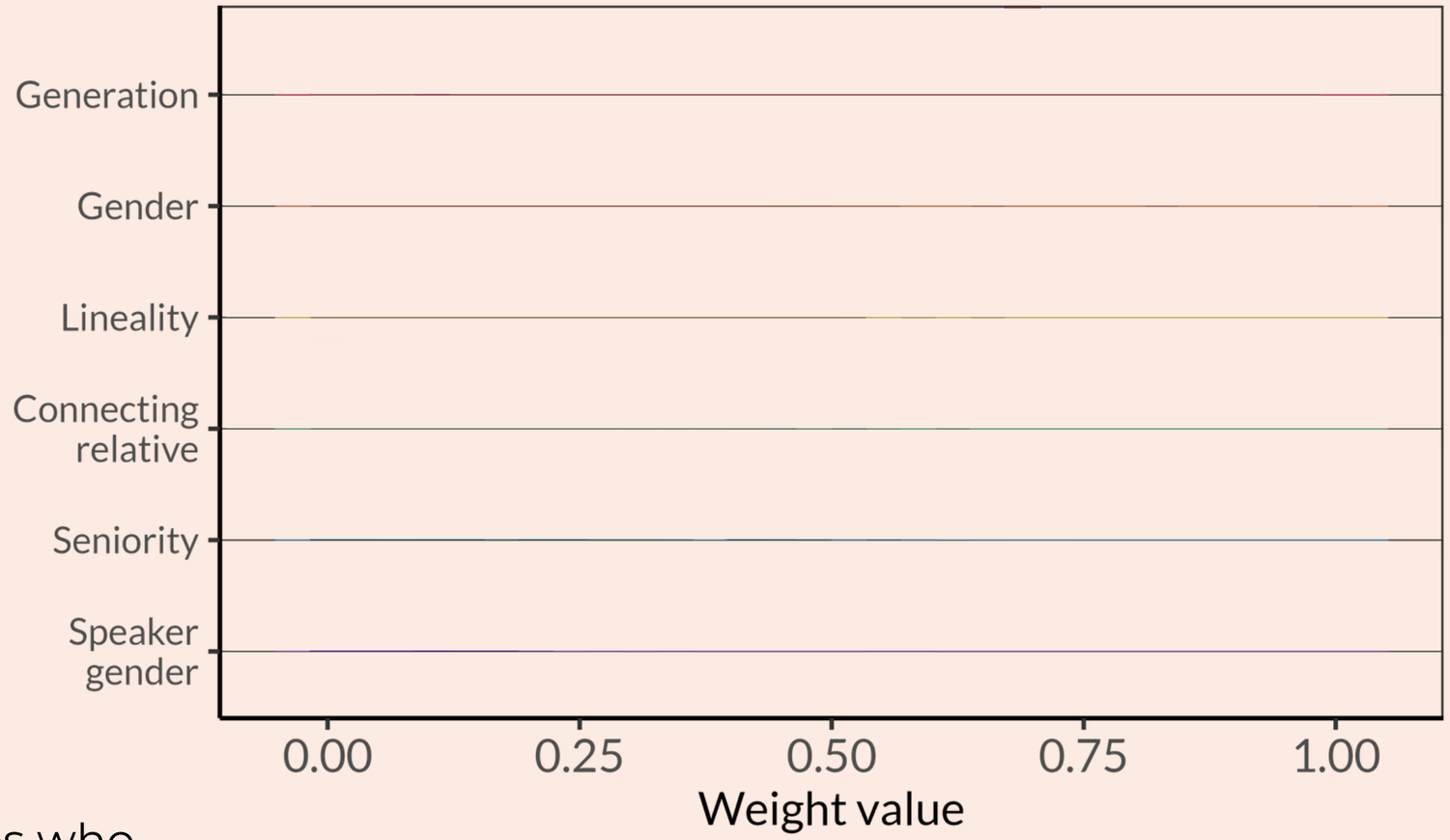


0.2 | 0.4 | 0.1 | 0.05 | 0.05 | 0.2

kind of rubbish correlation between form and meaning

And we find that some features are **more reliable cues to shared form across languages**!

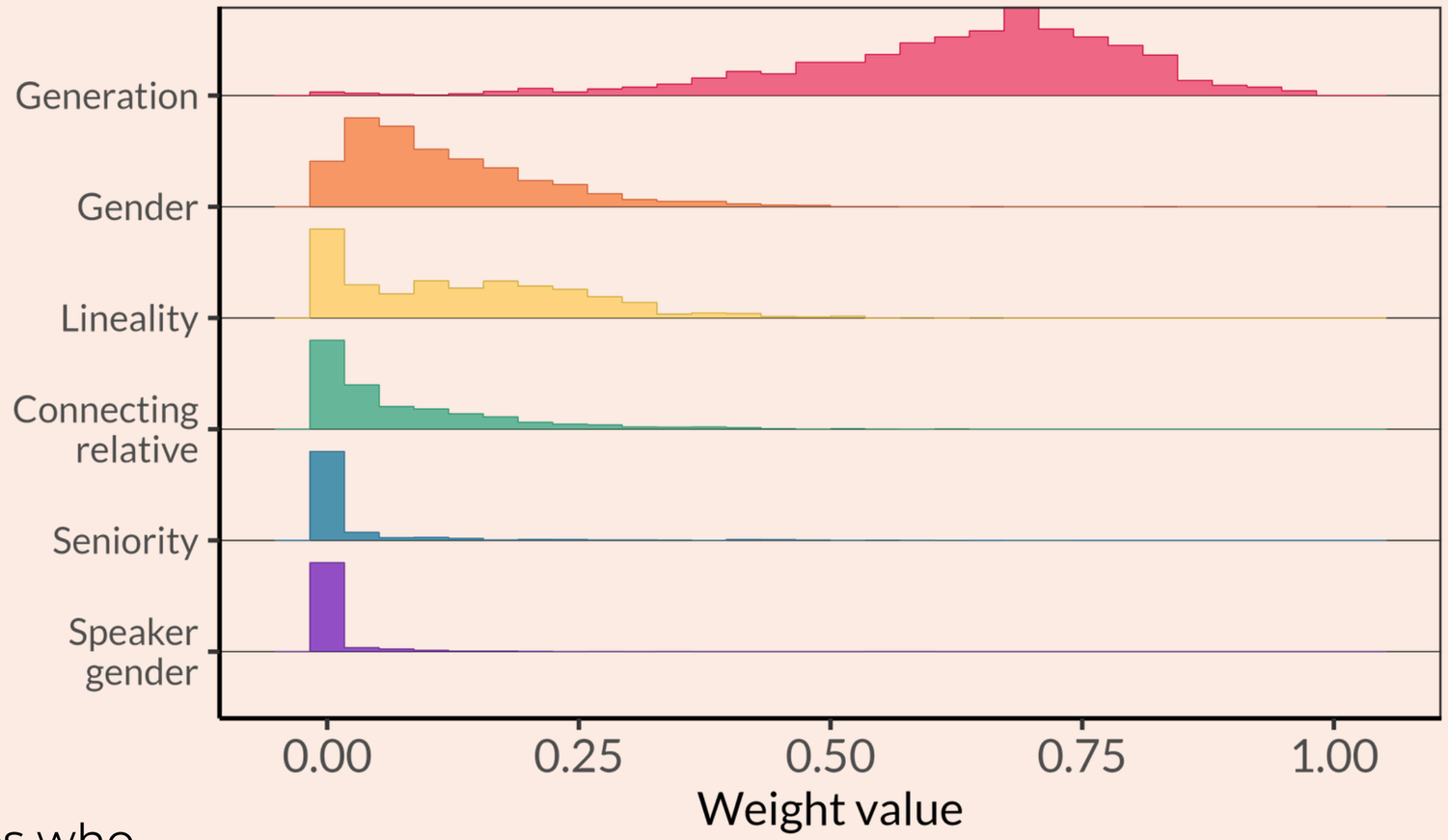
And we find that some features are **more reliable cues to shared form across languages**!



languages who distinguish **fewer** kin by this feature

languages who distinguish **more** kin by this feature

And we find that some features are **more reliable cues to shared form across languages**!



languages who distinguish **fewer** kin by this feature

languages who distinguish **more** kin by this feature

From this, we derived a **hierarchy of encoding** for kin term semantics.

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Connecting Relative

who connects you to this relative?

Seniority

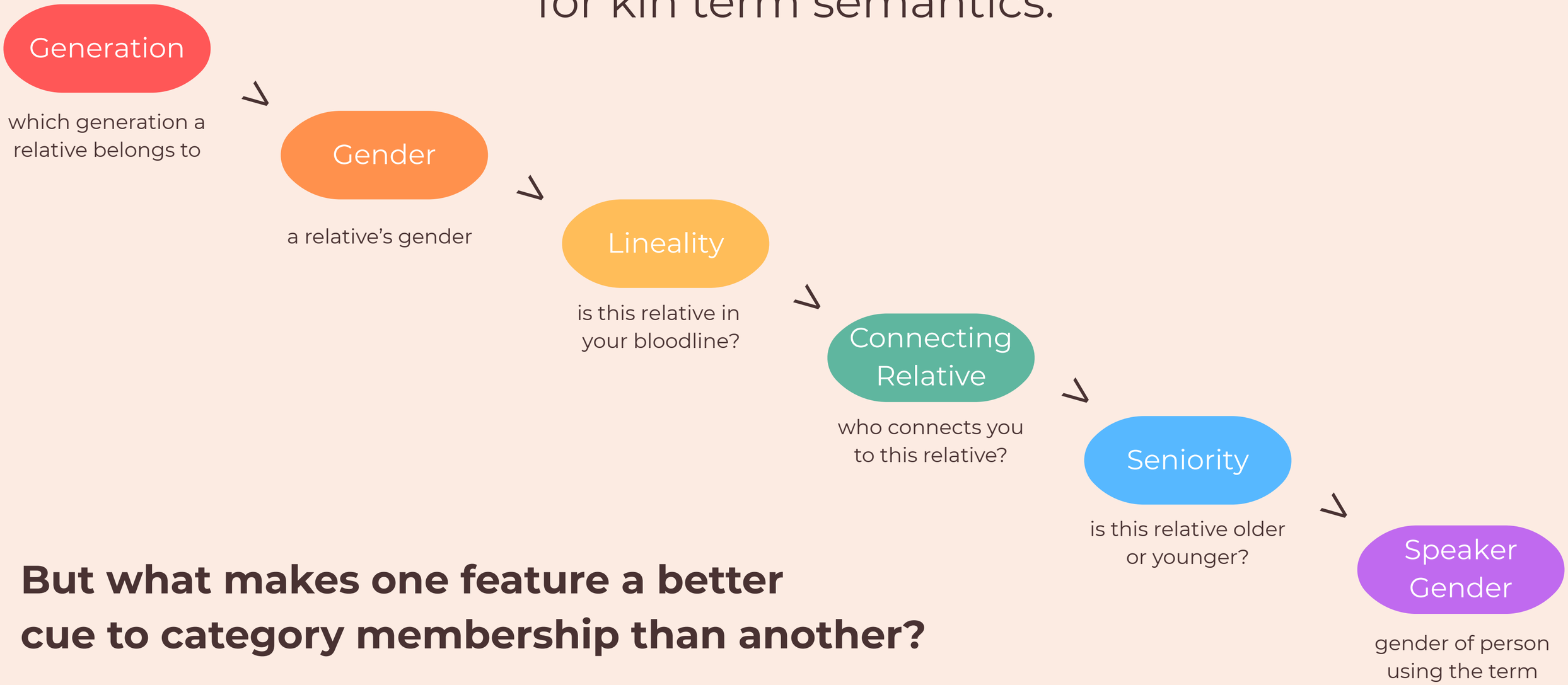
is this relative older or younger?

Speaker Gender

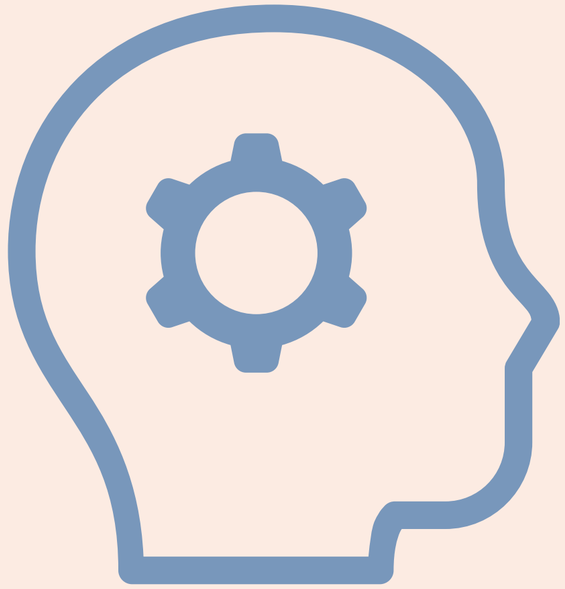
gender of person using the term

Maisy Hallam ✧ Centre for Language Evolution

From this, we derived a **hierarchy of encoding** for kin term semantics.



But what makes one feature a better cue to category membership than another?



Does this hierarchy align with people's preferences when classifying kin in an artificial language task?

Hypothesis:

When assigning labels to kin who share
competing semantic features,

people prefer to give the **same label**
to kin who share a feature that ranks higher on the hierarchy,

and a **different label** to kin who share a lower-ranked feature.

Most likely to
receive the same label

Generation

which generation a
relative belongs to

Gender

a relative's gender

Lineality

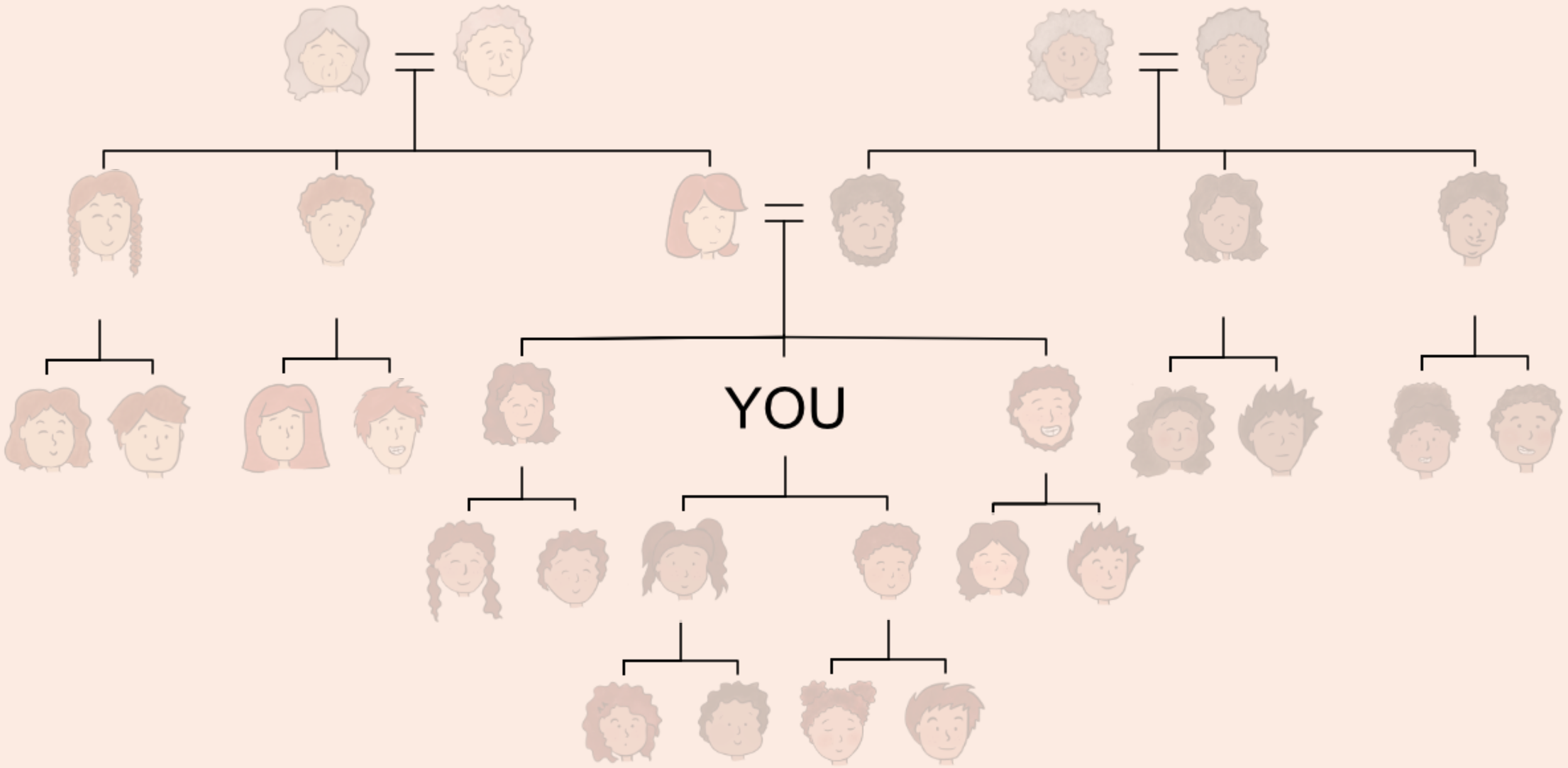
is this relative in
your bloodline?

Least likely to
receive the same label

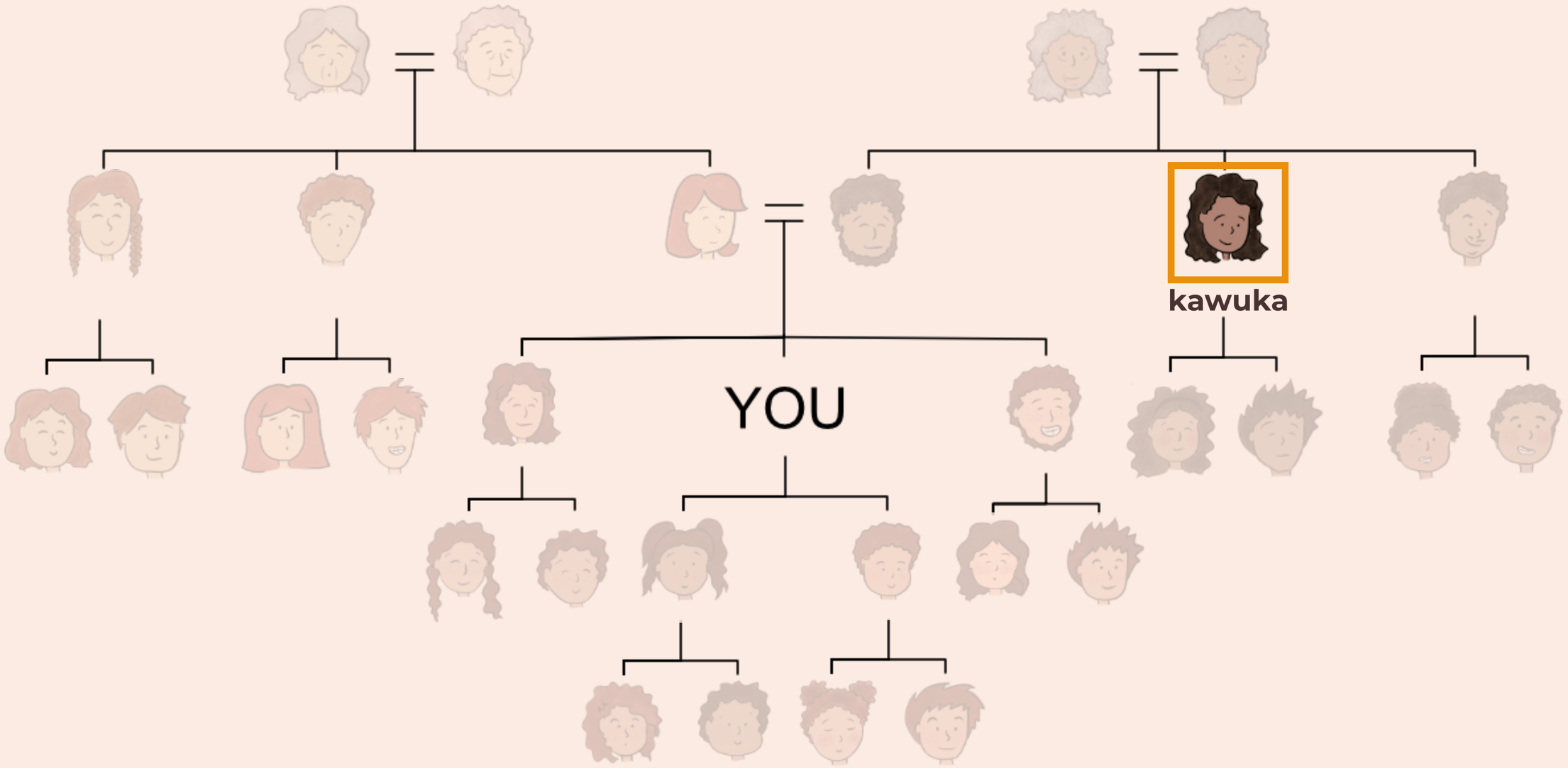
Connecting
Relative

who connects you
to this relative?

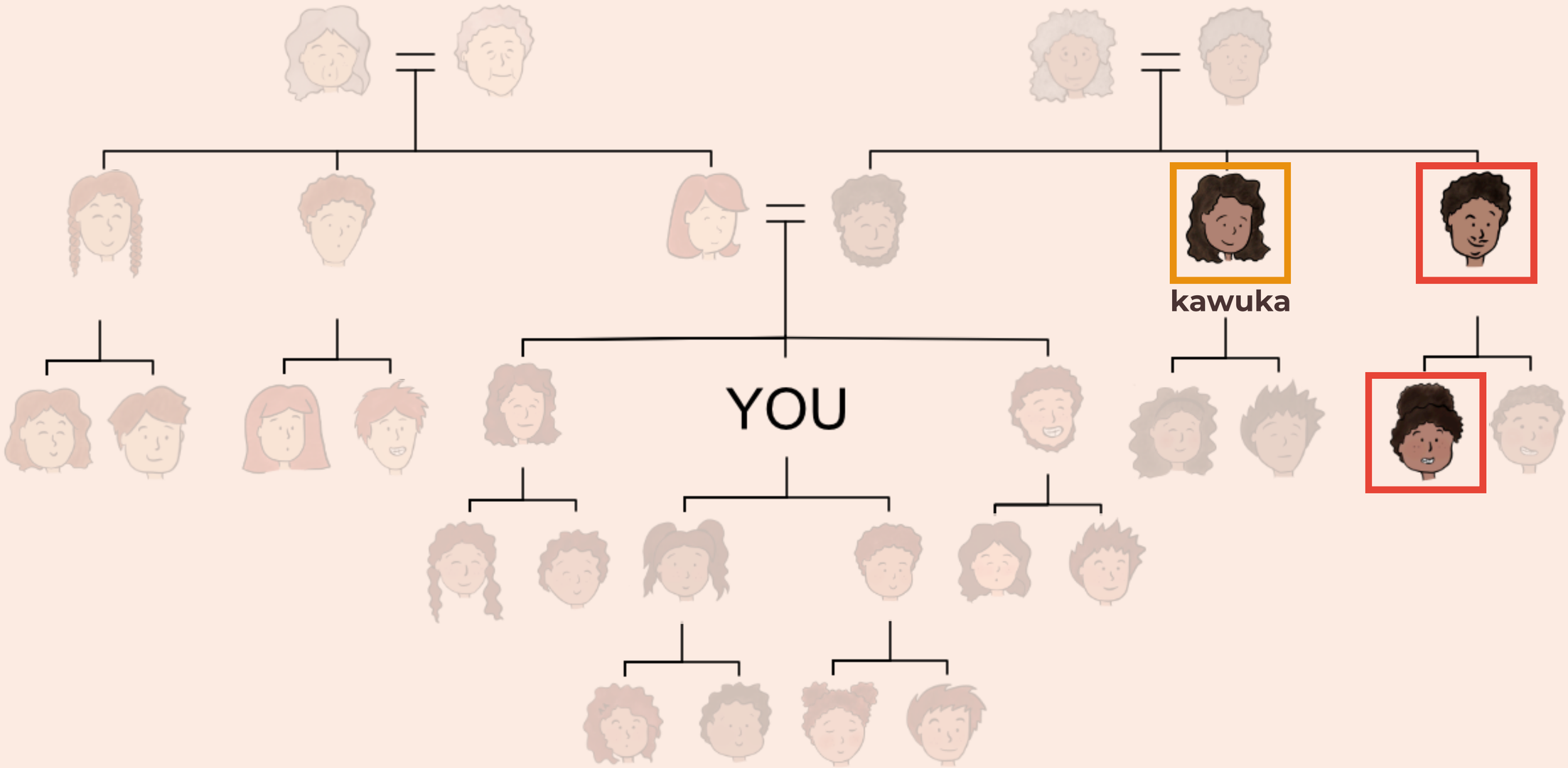
Maisy Hallam ✧ Centre for Language Evolution


YOU

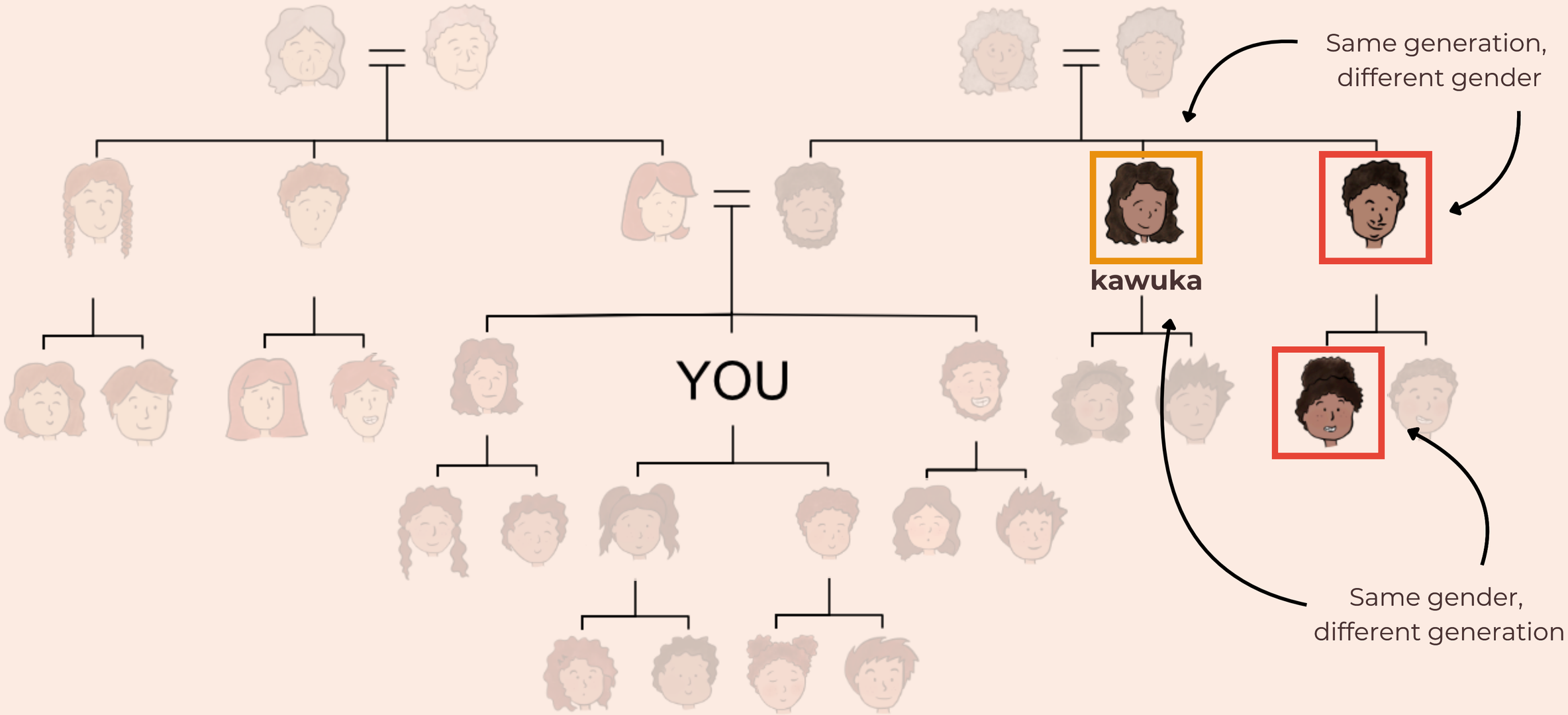
In this language, this family member is called your **kawuka**.



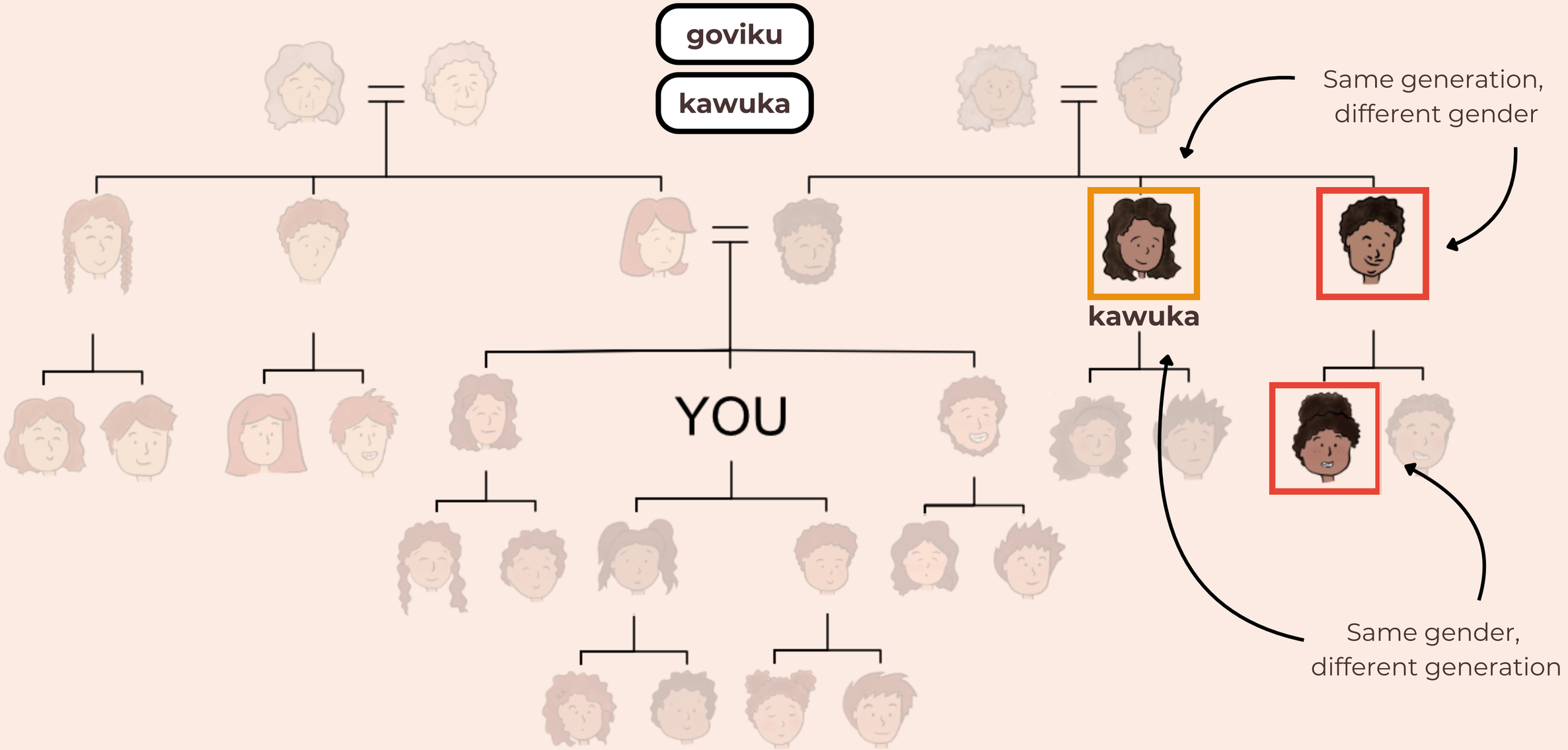
One of these two family members is also called your **kawuka**.



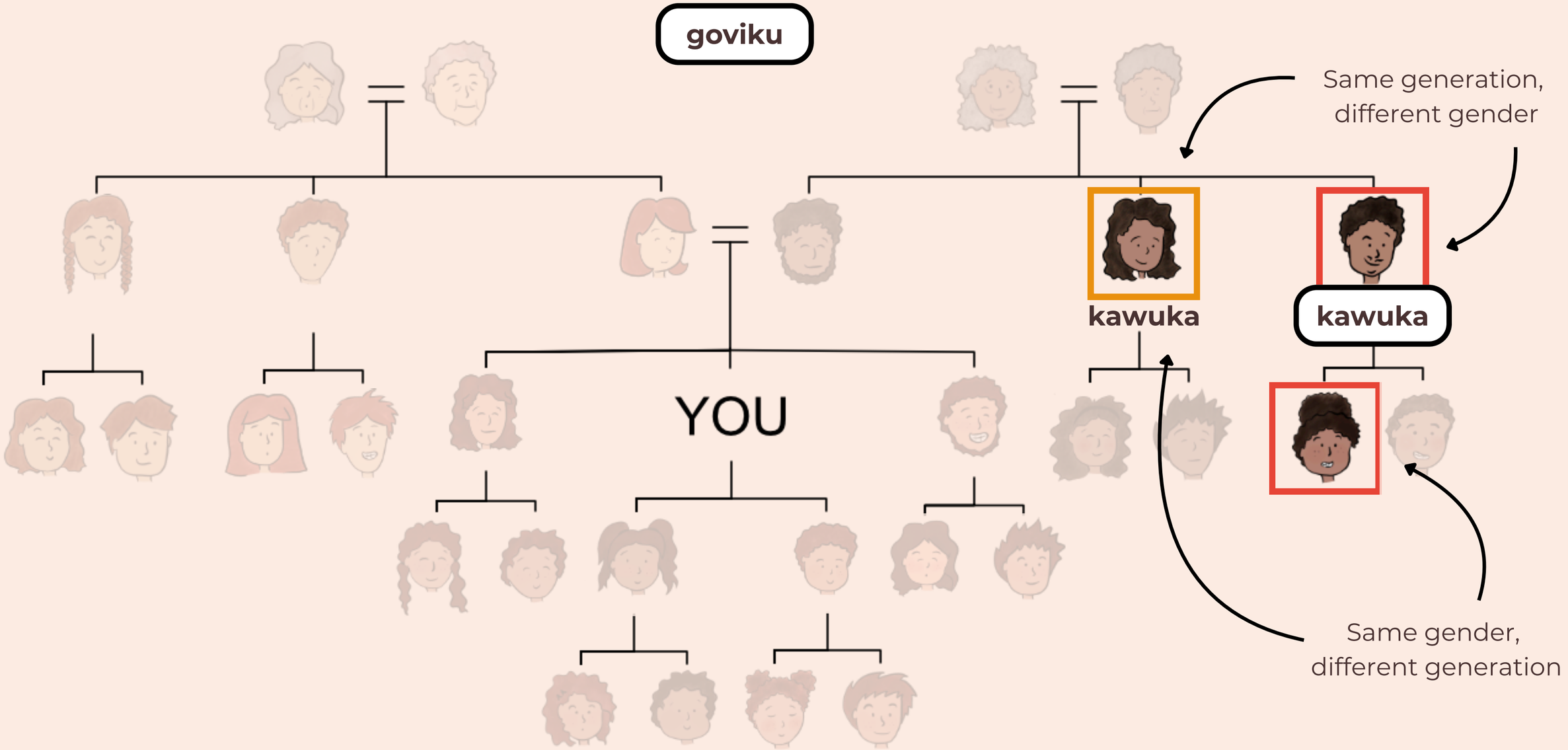
One of these two family members is also called your **kawuka**.



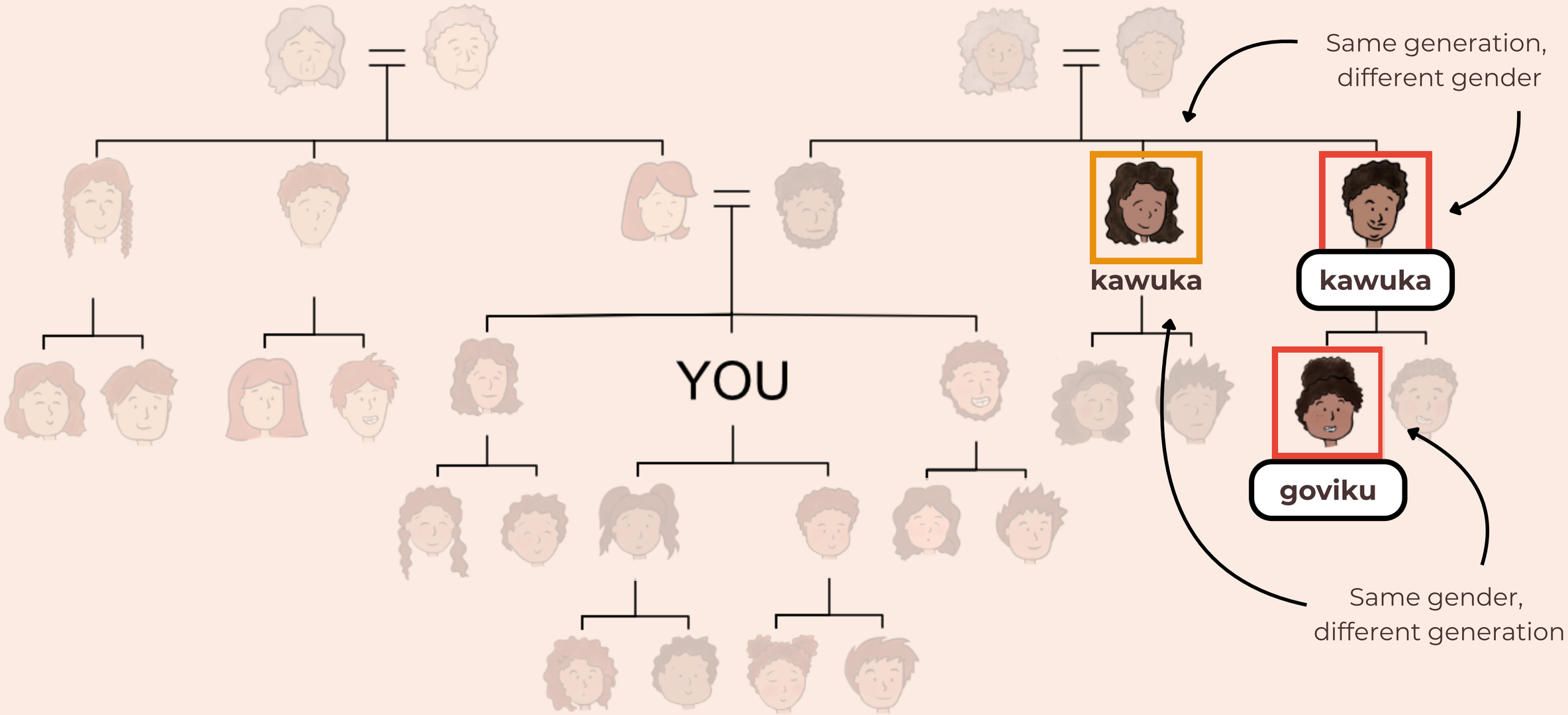
Choose a word for each family member by dragging and dropping the labels beneath them.

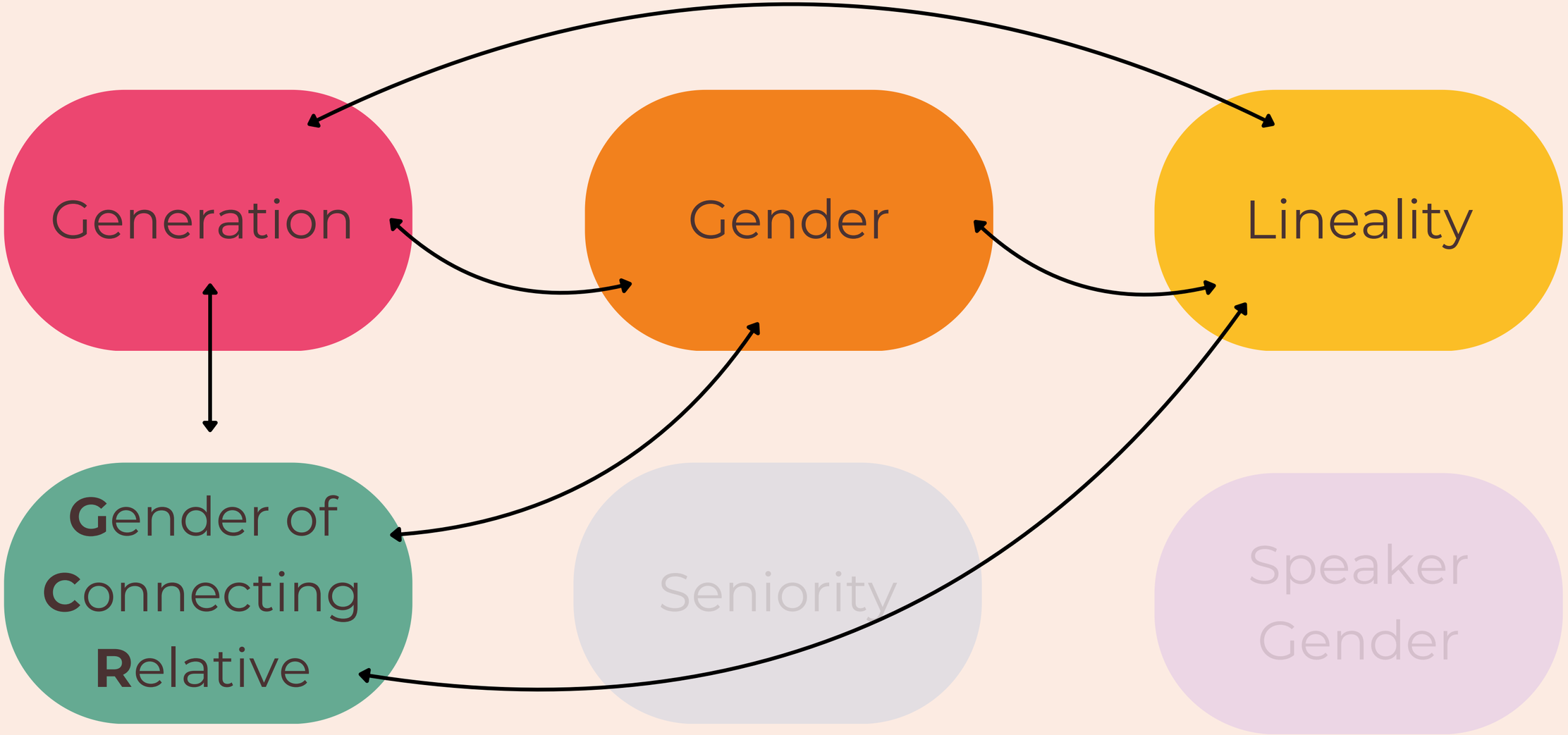


Choose a word for each family member by dragging and dropping the labels beneath them.

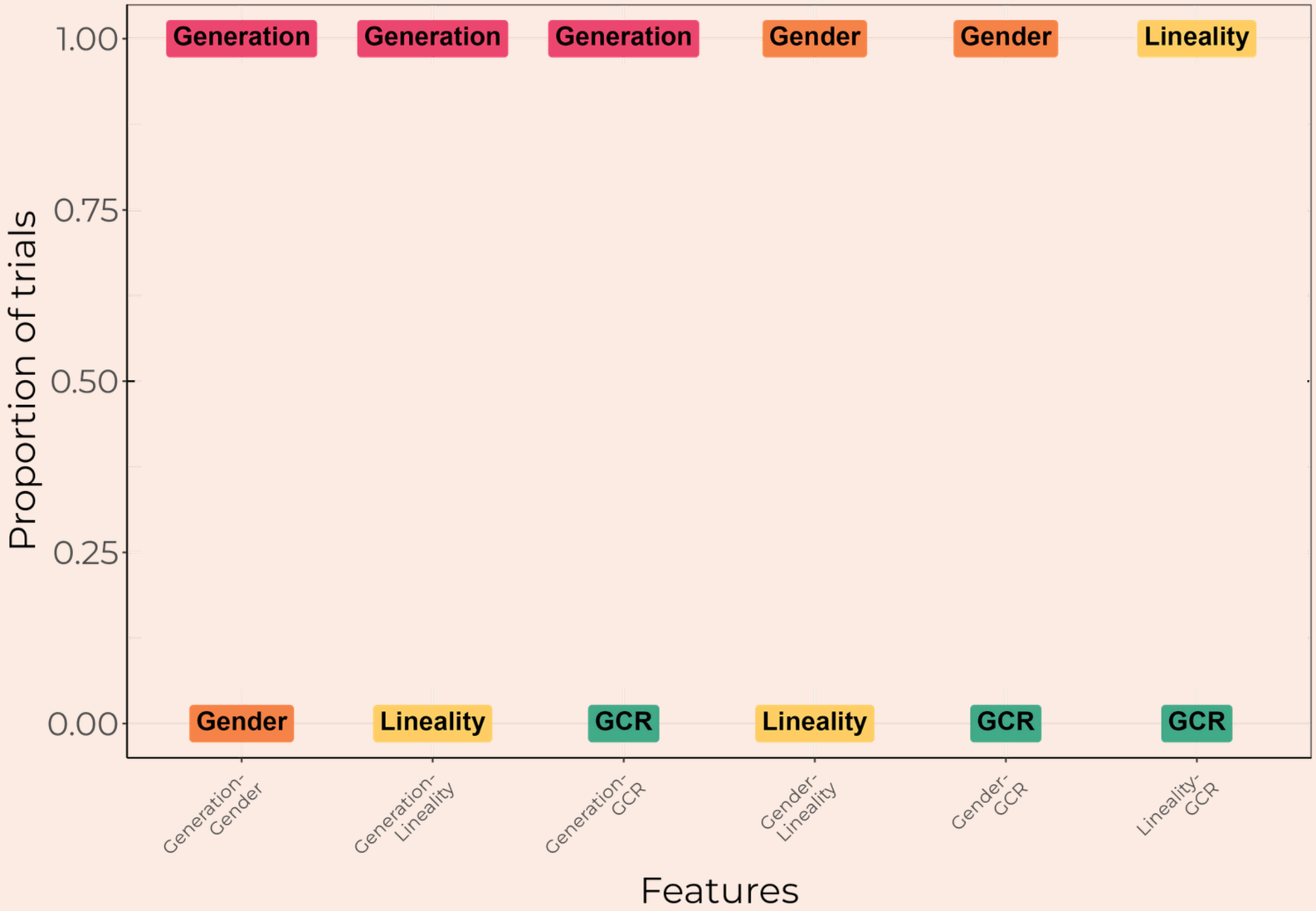


Choose a word for each family member by dragging and dropping the labels beneath them.

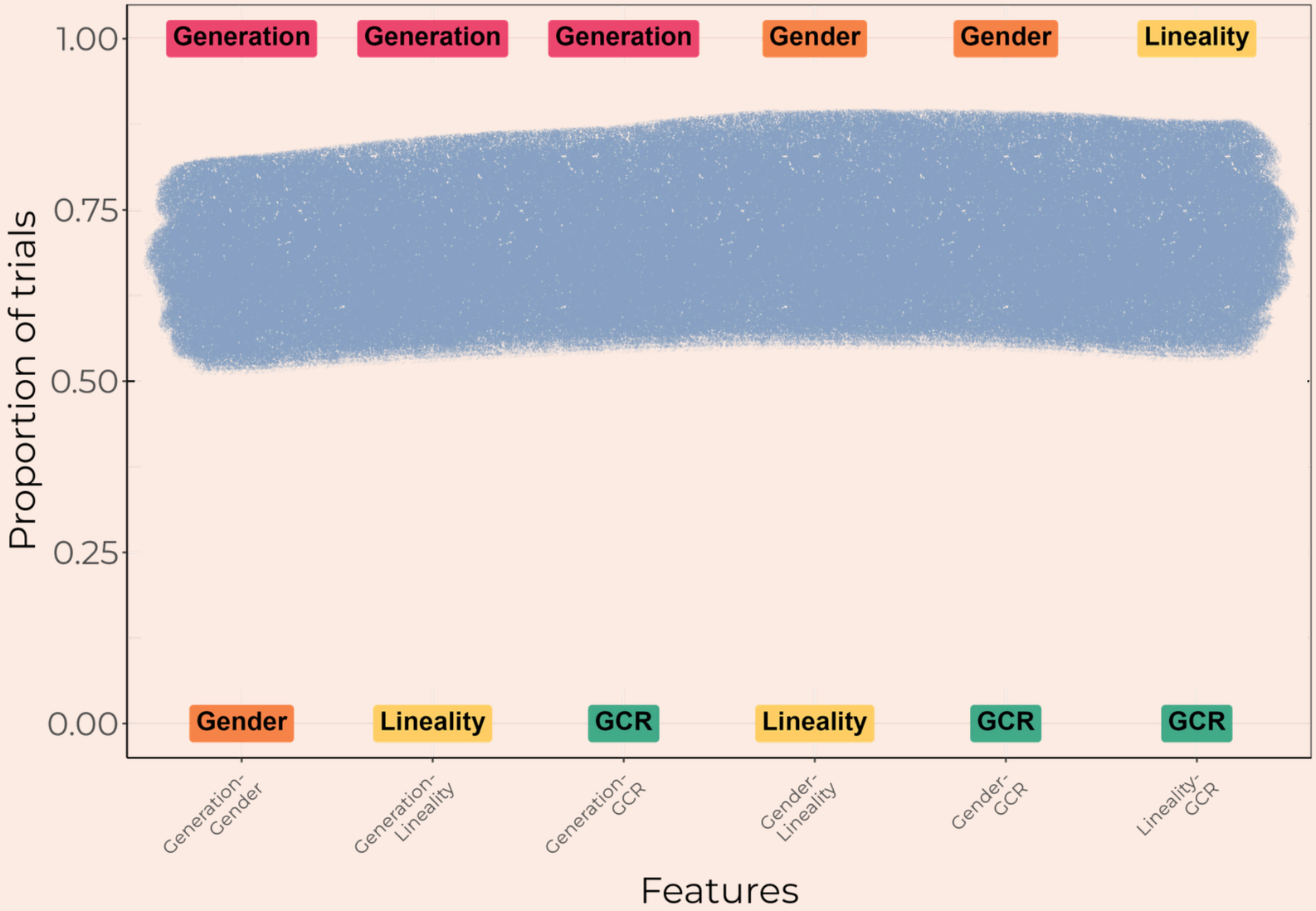




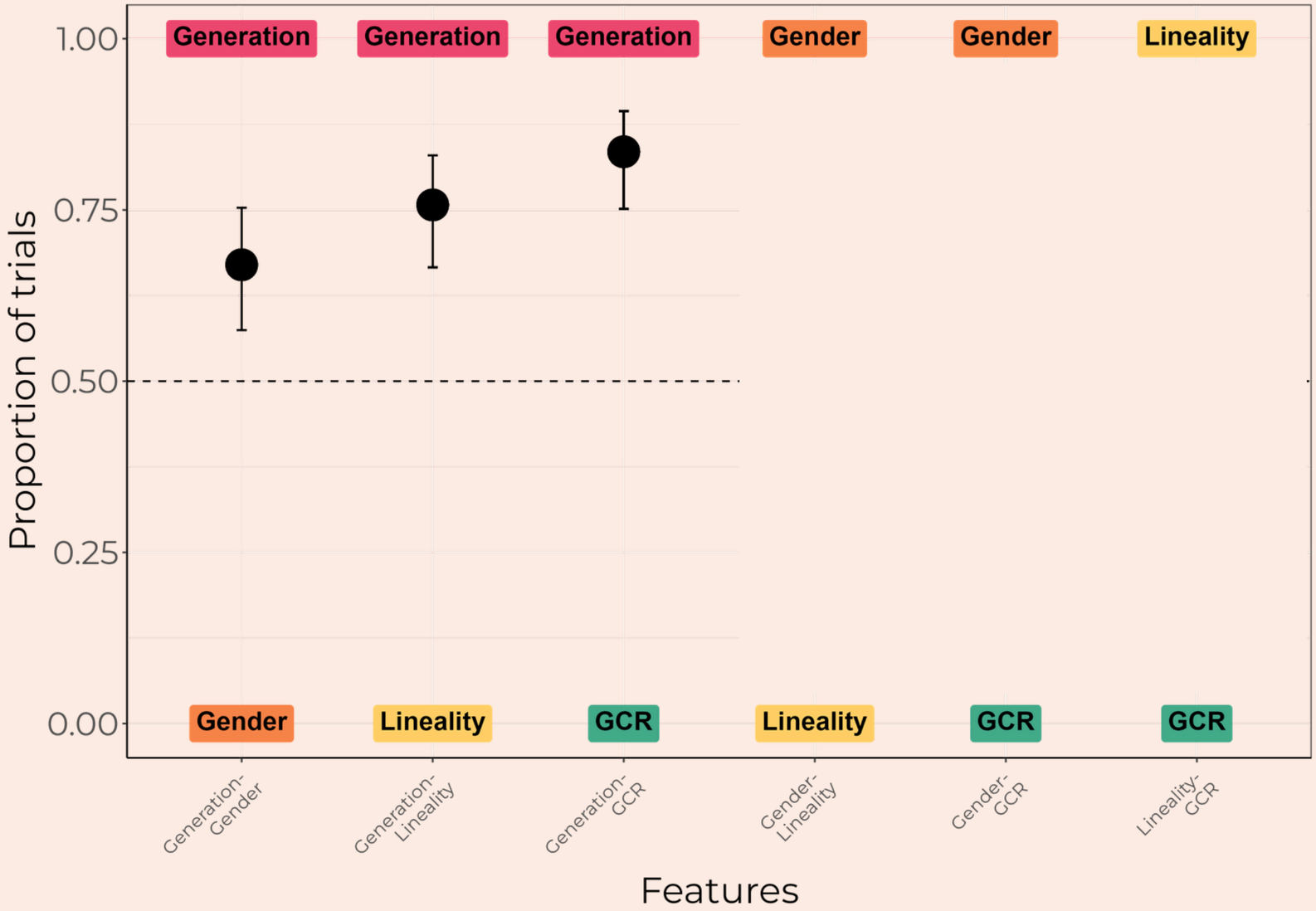
6 trials total



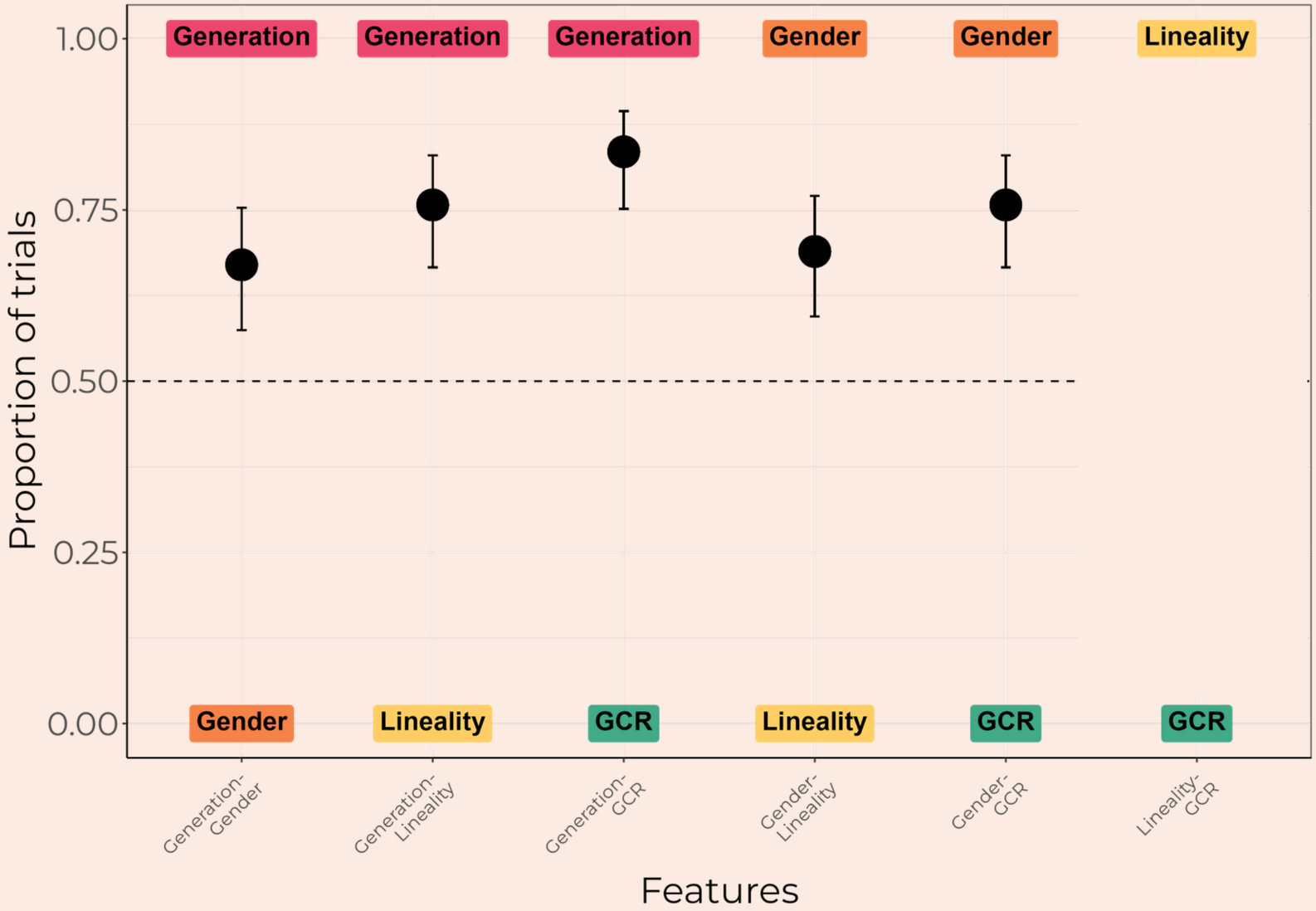
Participants prefer to categorise kin who share higher ranked features!



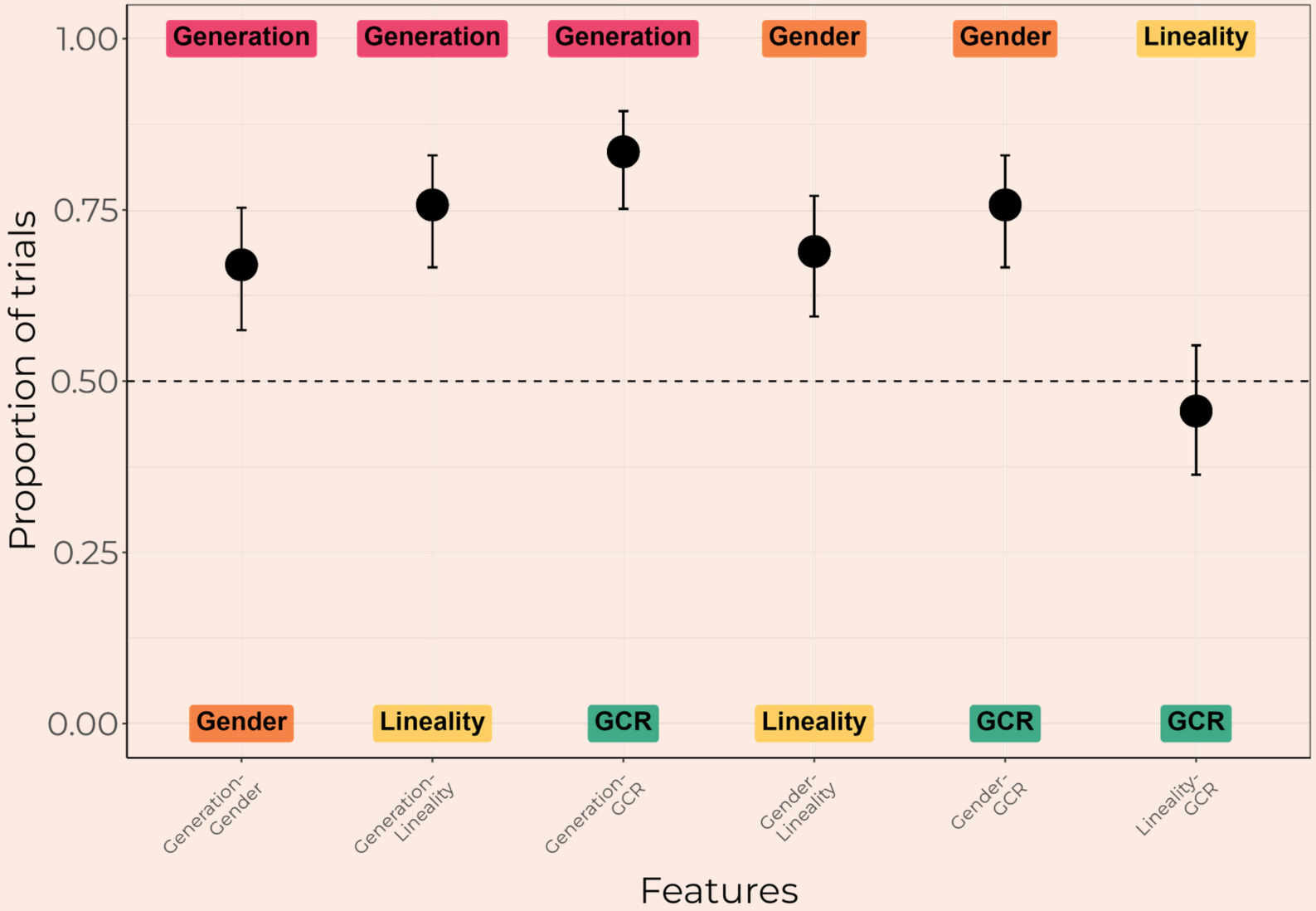
Participants prefer to categorise kin who share higher ranked features!

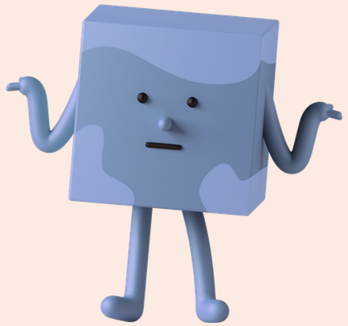


Participants prefer to categorise kin who share higher ranked features!



Participants prefer to categorise kin who share higher ranked features!





Participants prefer to categorise kin who share higher ranked features...?

Intrinsic properties of individuals

Relational properties

Intrinsic properties of individuals

Generation

Gender

Relational properties

Intrinsic properties of individuals

Generation

Gender

Relational properties

Lineality

Connecting Relative

Seniority

Intrinsic properties of individuals

Generation

Gender

Relational properties

Lineality

Connecting Relative

Seniority

?

Speaker Gender

Intrinsic properties of individuals

Generation

Gender

easy to identify!
easy to learn!

Relational properties

Lineality

Connecting Relative

Seniority

Speaker Gender

Asmuth and Gentner (2017)
Gentner and Kurtz (2005)
Gentner (2005)
Corral et al. (2017)

Intrinsic properties of individuals

Generation

Gender

easy to identify!
easy to learn!

Relational properties

hard to identify!
even harder to learn!

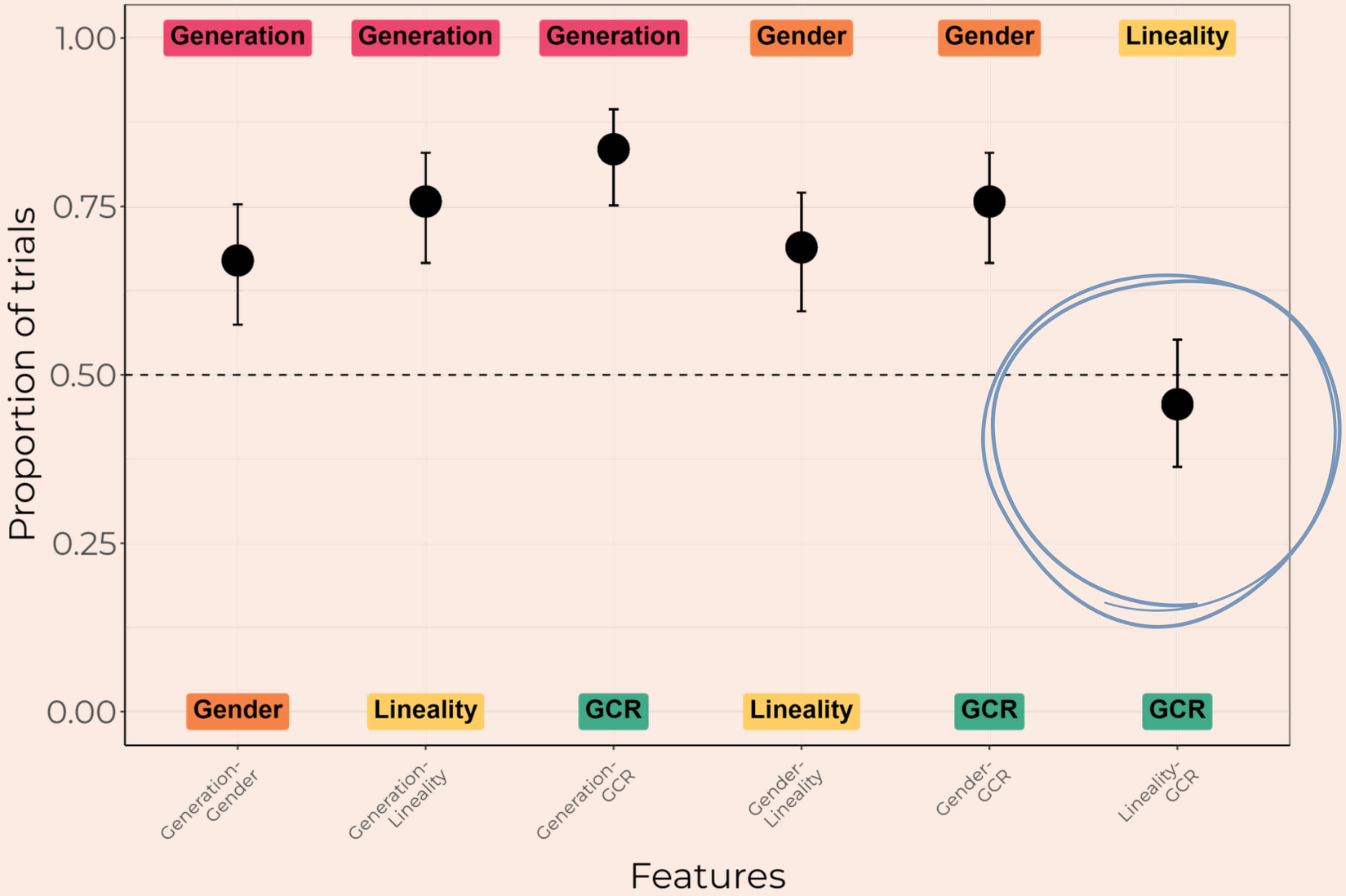
Lineality

Connecting Relative

Seniority

Speaker Gender

Asmuth and Gentner (2017)
Gentner and Kurtz (2005)
Gentner (2005)
Corral et al. (2017)



Which might explain why our participants struggle so much on this trial.

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Connecting Relative

who connects you to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

Maisy Hallam ✧ Centre for Language Evolution

Higher ranked features are good cues to category membership because they are **easy to identify and learn...**

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

is this relative in your bloodline?

Connecting Relative

who connects you to this relative?

Seniority

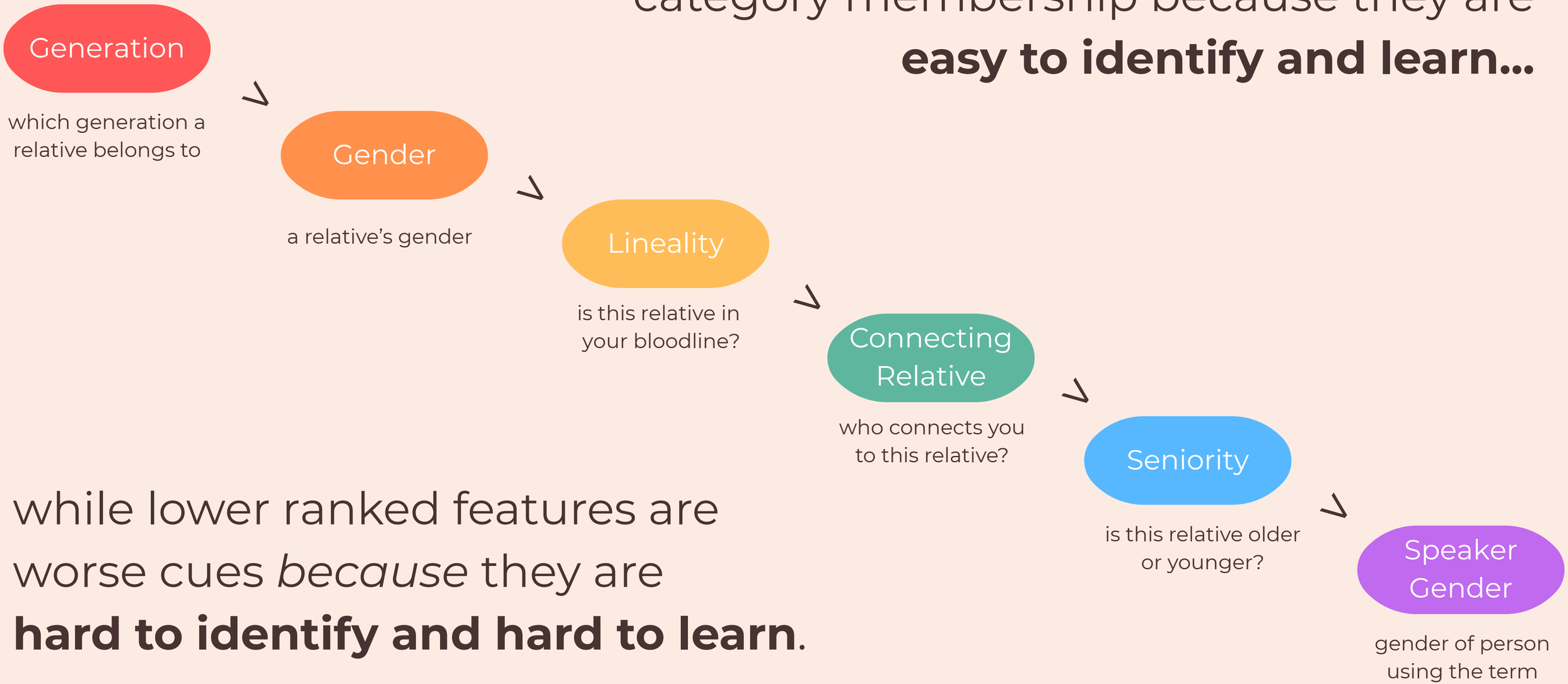
is this relative older or younger?

Speaker Gender

gender of person using the term

Maisy Hallam ✧ Centre for Language Evolution

Higher ranked features are good cues to category membership because they are **easy to identify and learn...**



while lower ranked features are worse cues *because* they are **hard to identify and hard to learn**.

Maisy Hallam ✧ Centre for Language Evolution

ASK ME ABOUT COMMUNICATIVE NEED (if you like)

Higher ranked features are good cues to category membership because they are **easy to identify and learn...**

Generation

which generation a relative belongs to

Gender

a relative's gender

Lineality

this relative in your bloodline?

Connecting Relative

to this relative?

Seniority

is this relative older or younger?

Speaker Gender

gender of person using the term

while lower ranked features are worse cues *because* they are **hard to identify and hard to learn**.

Maisy Hallam ✧ Centre for Language Evolution

Thanks for coming!

Maisy Hallam

Individual preferences during classification
align with kinship semantics across languages

https://maisyhallam.github.io/

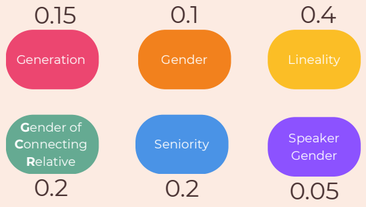
maisy.hallam@ed.ac.uk

extra extra, read all about it!
(CogSci Proceedings 2025)

model.

Our model **optimises the correlation** between
similarity in form and similarity in meaning through gradient ascent.

1



2

3

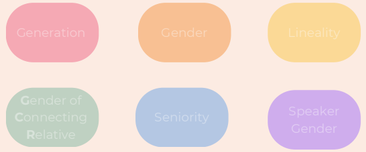
4

0.15
Generation
0.1
Gender
0.4
Lineality
Gender of Connecting Relative
0.2
Seniority
0.2
Speaker Gender
0.05

First, randomly assign a weight to each semantic feature setting.

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.

1



2

3

4

0.15*1

Generation

0.1*-1

Gender

0.4*1

Lineality

Gender of Connecting Relative

0.2*0

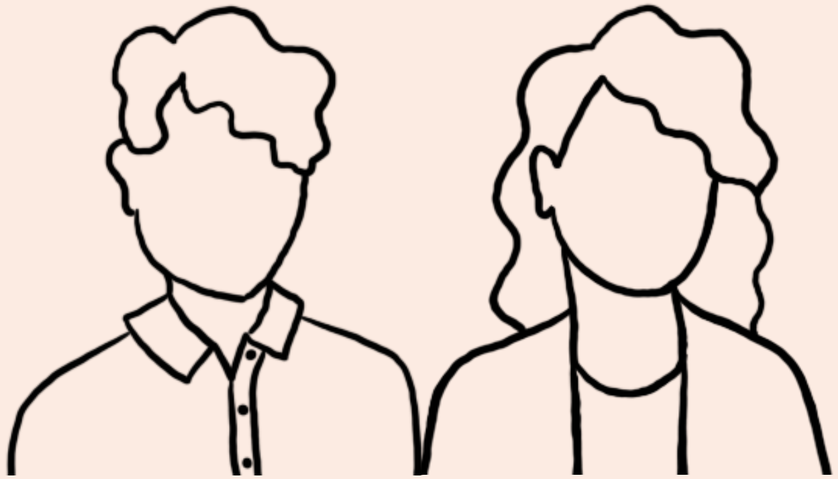
Seniority

0.2*1

Speaker Gender

0.05*-1

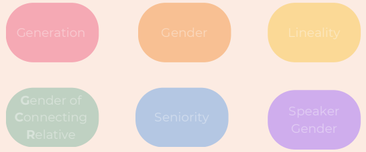
= 0.6



For each pair of kin, calculate the **weighted sum of features shared**, and the proportion of their kin term shared.

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.

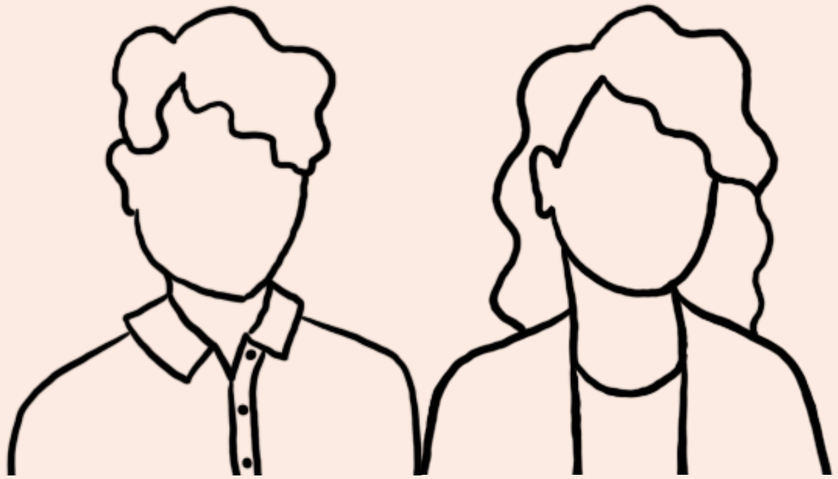
1



2

3

4



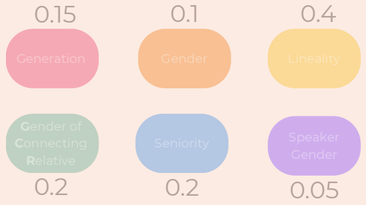
‘bro**th**e**r**’

‘sis**t**e**r**’

For each pair of kin, calculate the weighted proportion of features shared,
and the **proportion of their kin term shared**.

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.

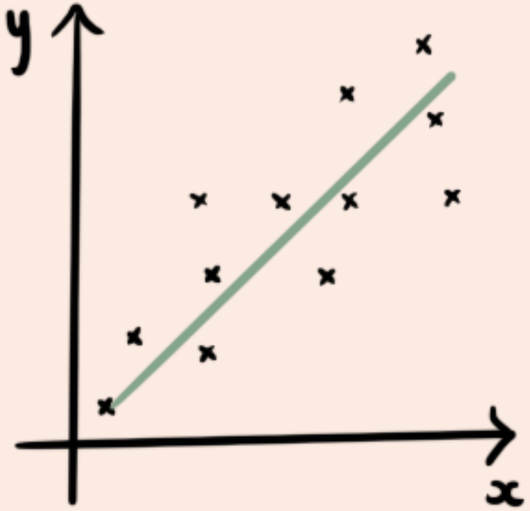
1



2

3

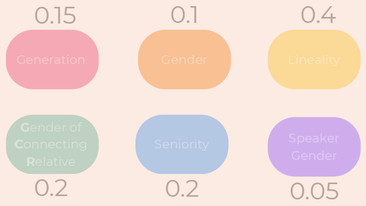
4



Over all pairs, calculate the correlation between
the weighted sum of shared features and proportion of shared forms

Our model **optimises the correlation** between
similarity in form and similarity in meaning through gradient ascent.

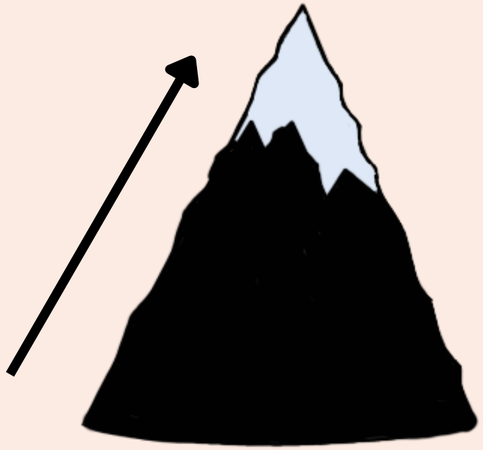
1



2

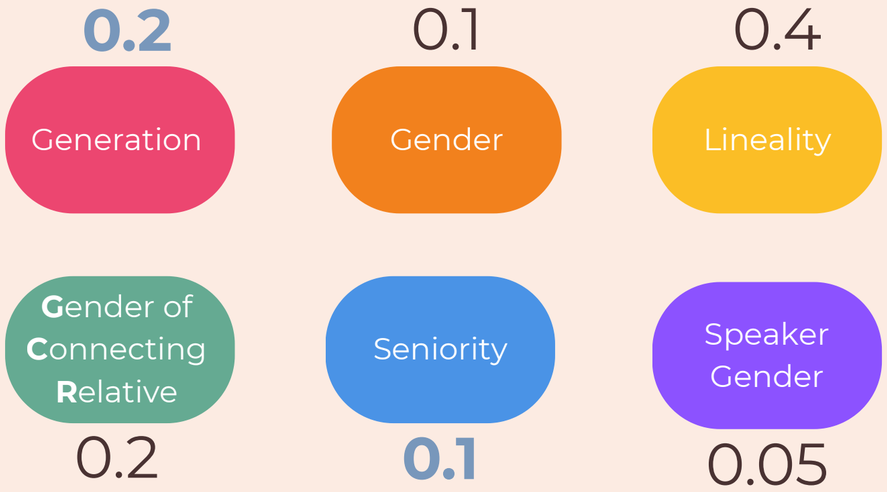
3

4



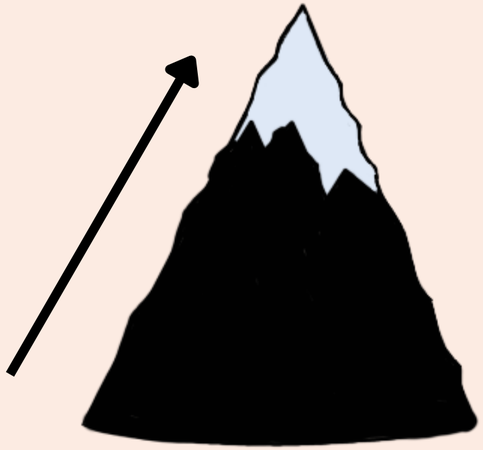
Could a change to the weights improve the correlation?
If yes...

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.



Make a small adjustment to the weights, and go again!

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.

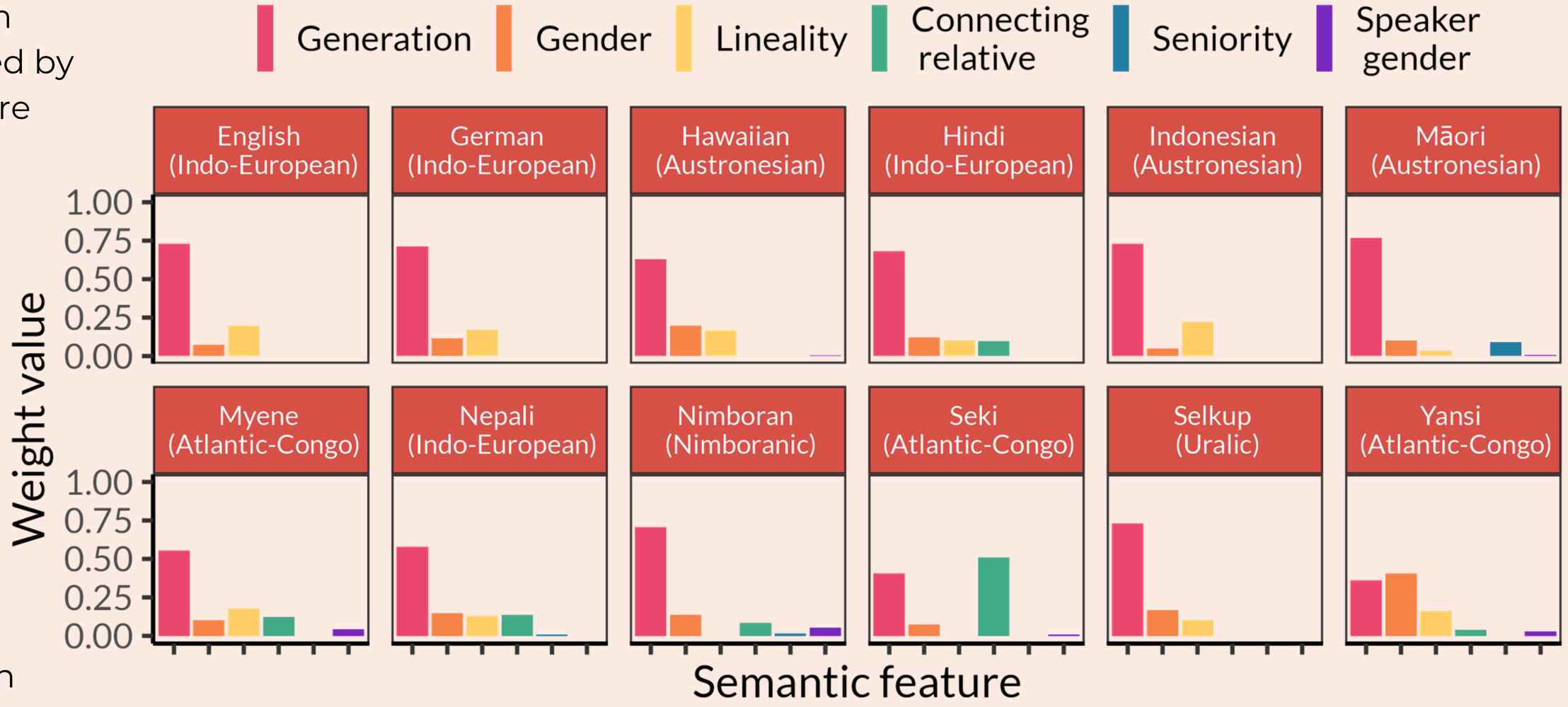


Could a change to the weights improve the correlation? If not, stop.

Our model **optimises the correlation** between similarity in form and similarity in meaning through gradient ascent.

further model results.

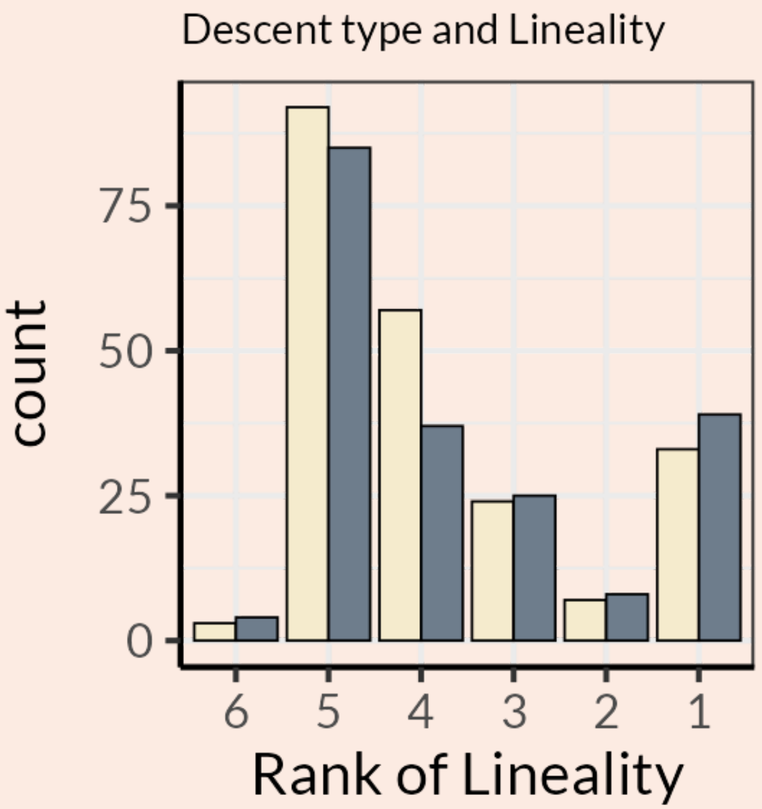
more kin distinguished by this feature

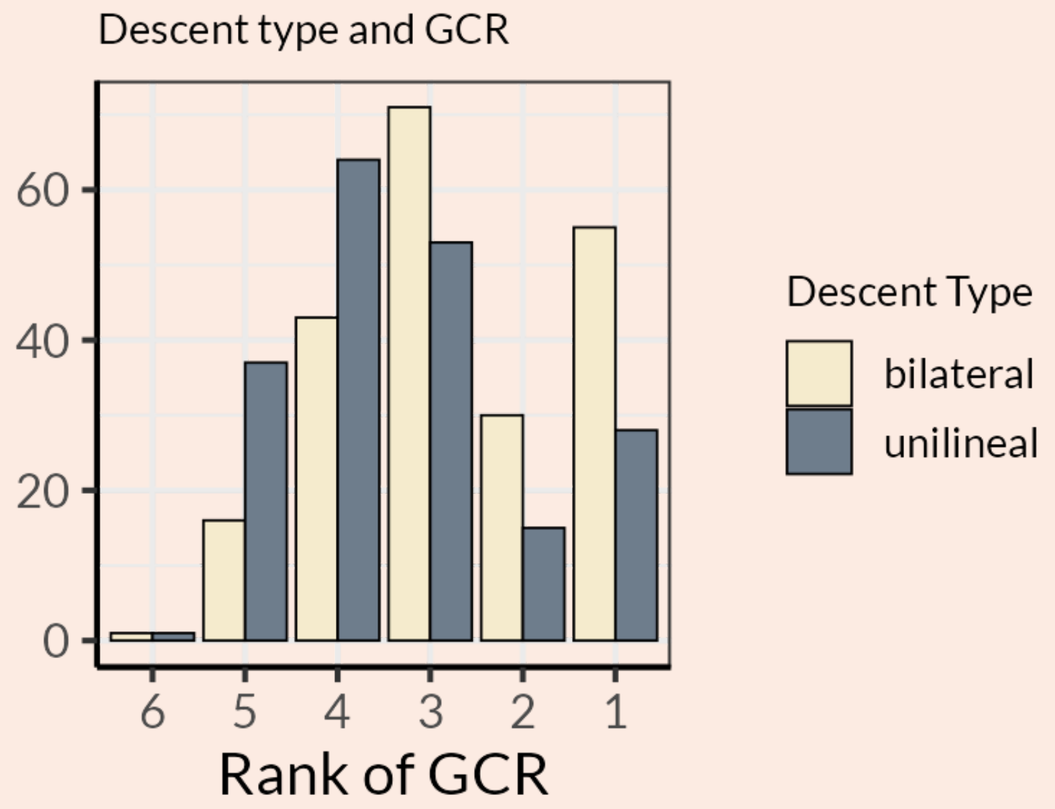


fewer kin distinguished by this feature

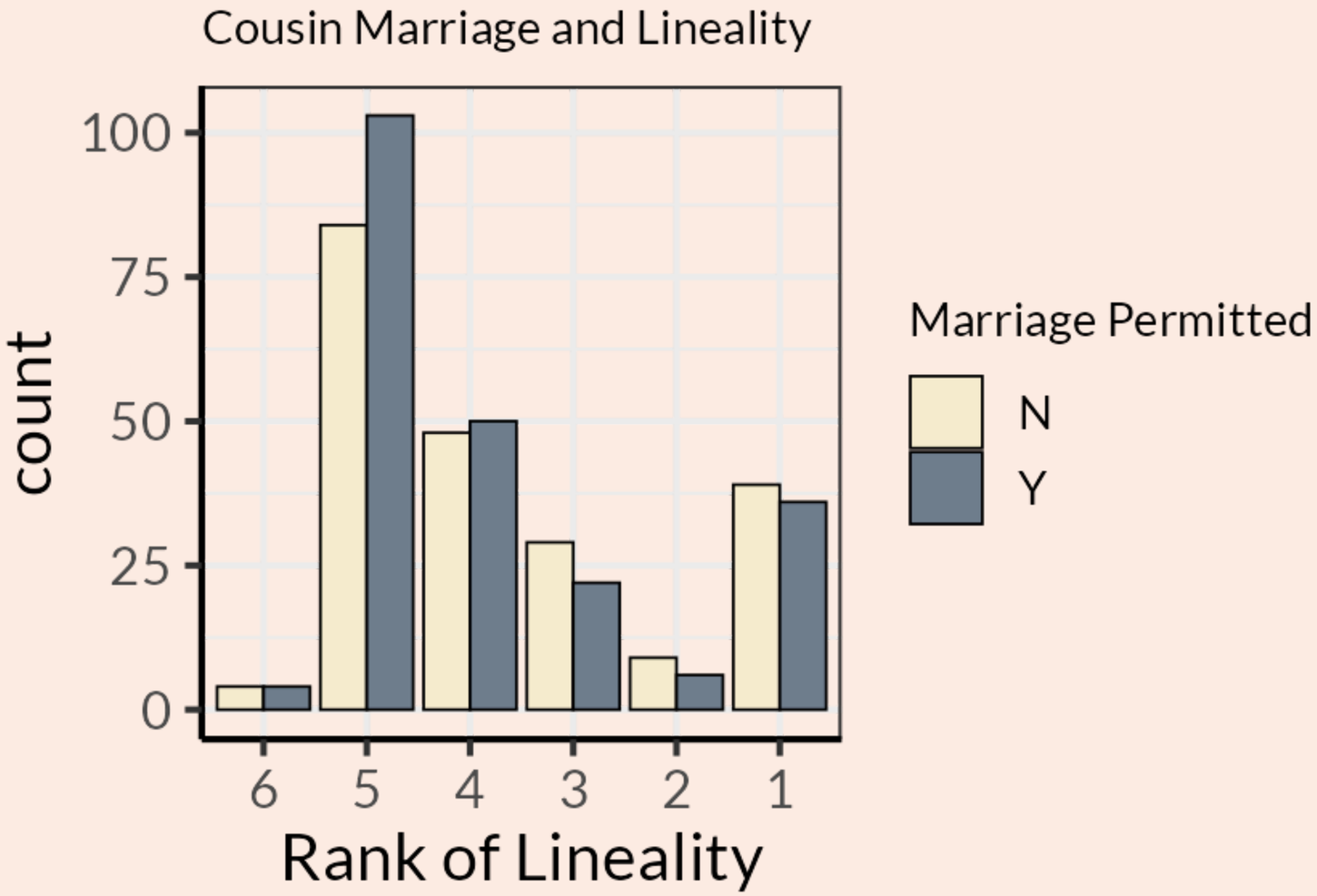
Sure enough, different languages encode different semantics with their kinship terminology.

social practice.

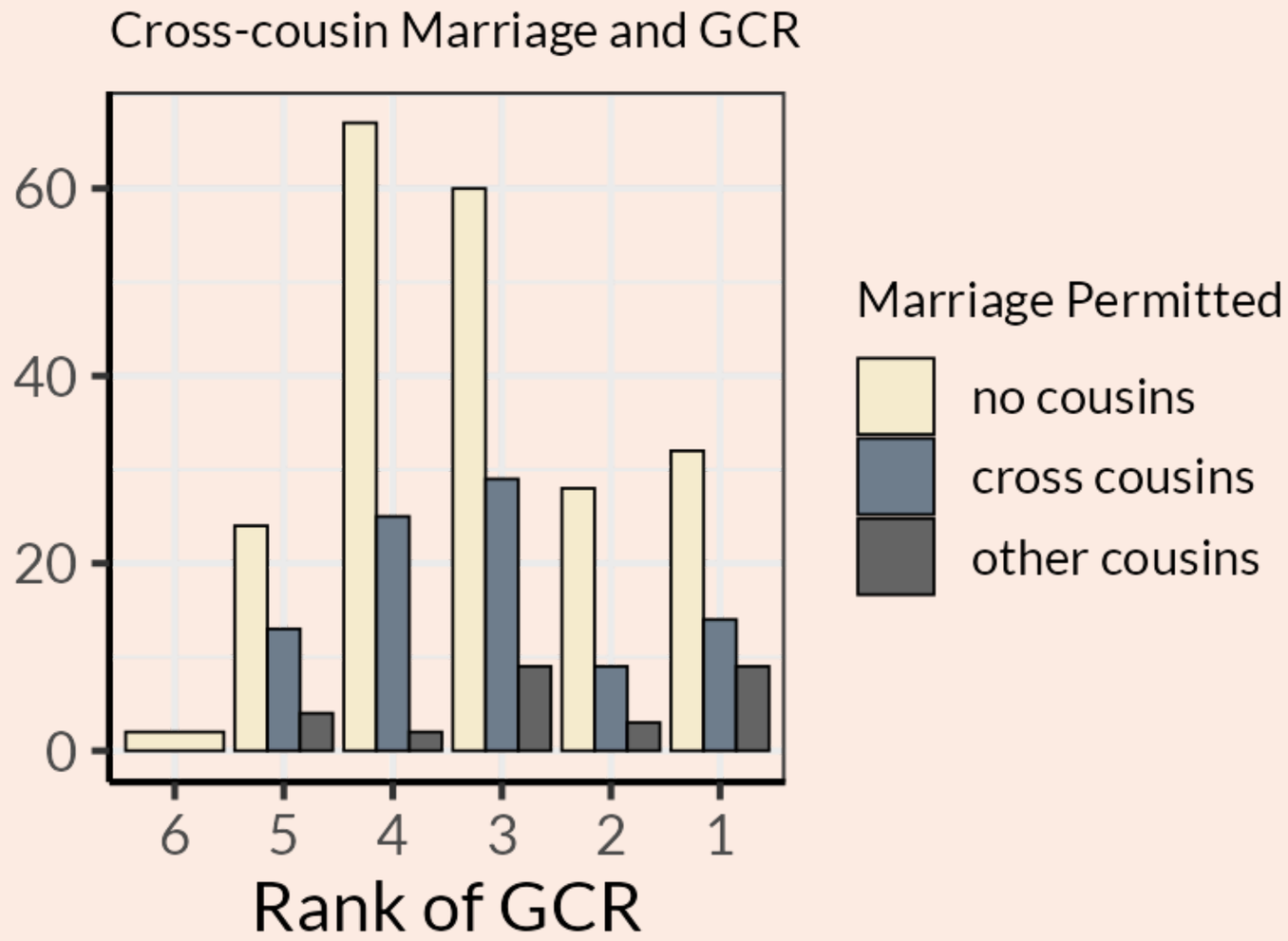




Societies with **bilateral descent** encode more Lineality distinctions than societies with **unilineal descent,** but fewer GCR distinctions!



Societies who **forbid cousin marriage** do not encode significantly more Lineality distinctions



Societies who **permit cross-cousin marriage** do not encode significantly more GCR distinctions.