

HIDDEN BIAS: IMPLICIT, EXPLICIT & DUAL PROCESS MODELS



Dr Katie Sullivan

THIS WEEK...

The Two
Systems of
Thinking

Implicit
Attitudes

Explicit
Attitudes

Foundational
Dual Process
Models

Modern Dual
Attitude
Models

Research
Methods

Behavioural
Predictions

Future
Research

WHY DOESN'T ALL PREJUDICE LOOK THE SAME?

Person A

- Openly expresses racist beliefs
- Actively avoids certain groups

Person B

- Strongly supports equality
- Rejects racism and discrimination
- Nevertheless, shows subtle bias in decisions

Who is prejudiced?

TWO SYSTEMS OF THINKING

System 1

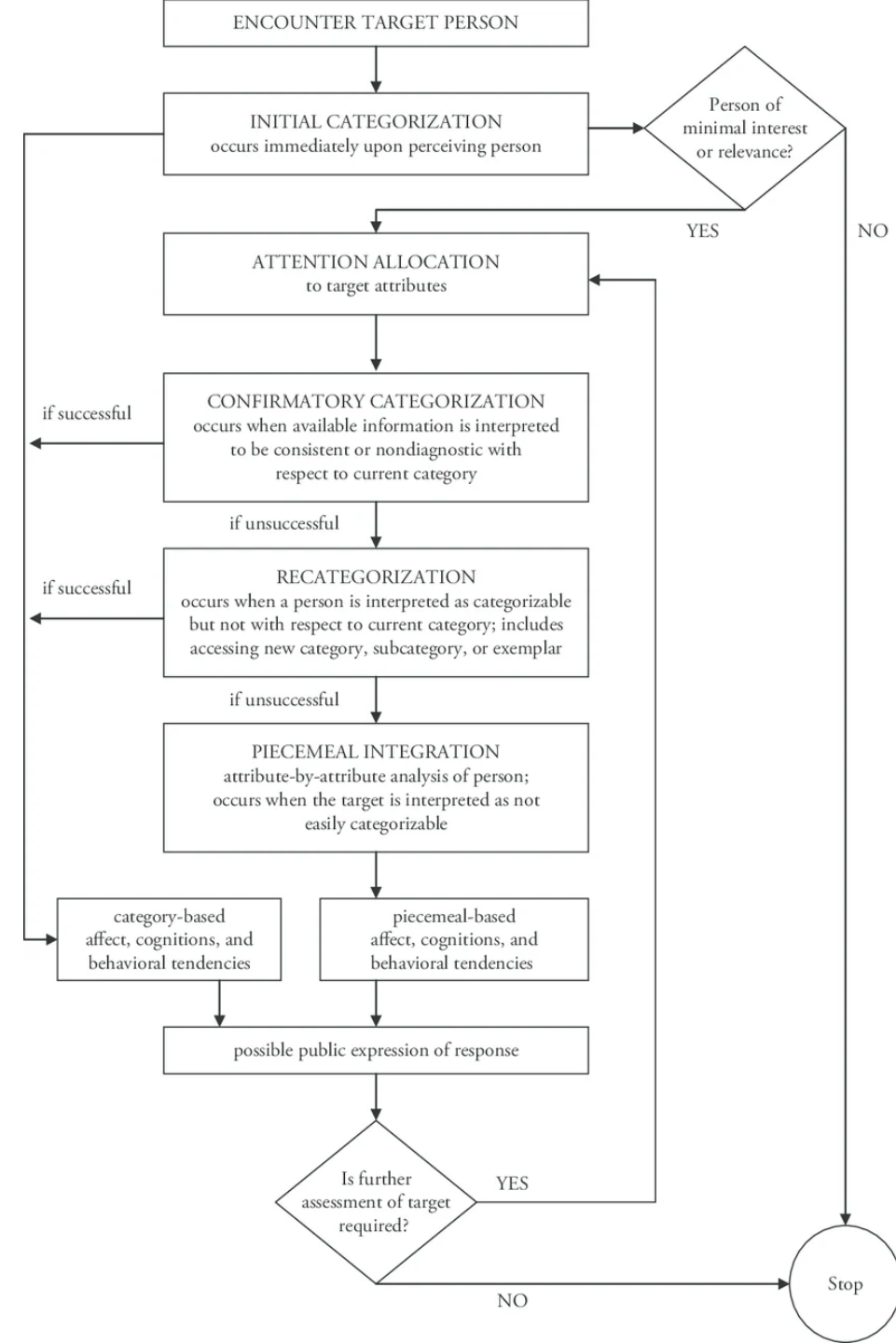
- Fast
- Automatic
- Effortless
- Intuitive
- Often unconscious

System 2

- Slow
- Deliberate
- Effortful
- Reflective
- Conscious

Can prejudice emerge from both systems?

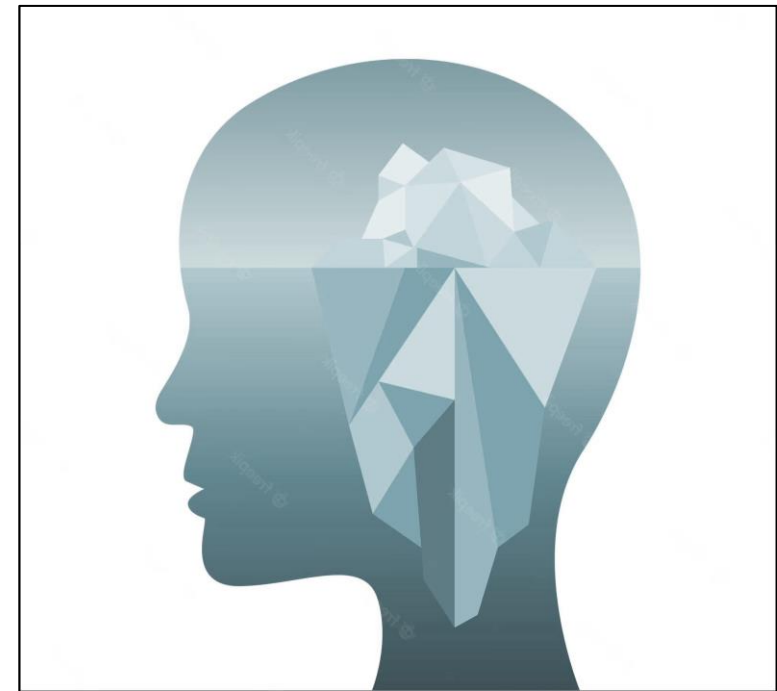
THE CONTINUUM MODEL OF IMPRESSION FORMATION (FISKE & NEUBERG, 1990)



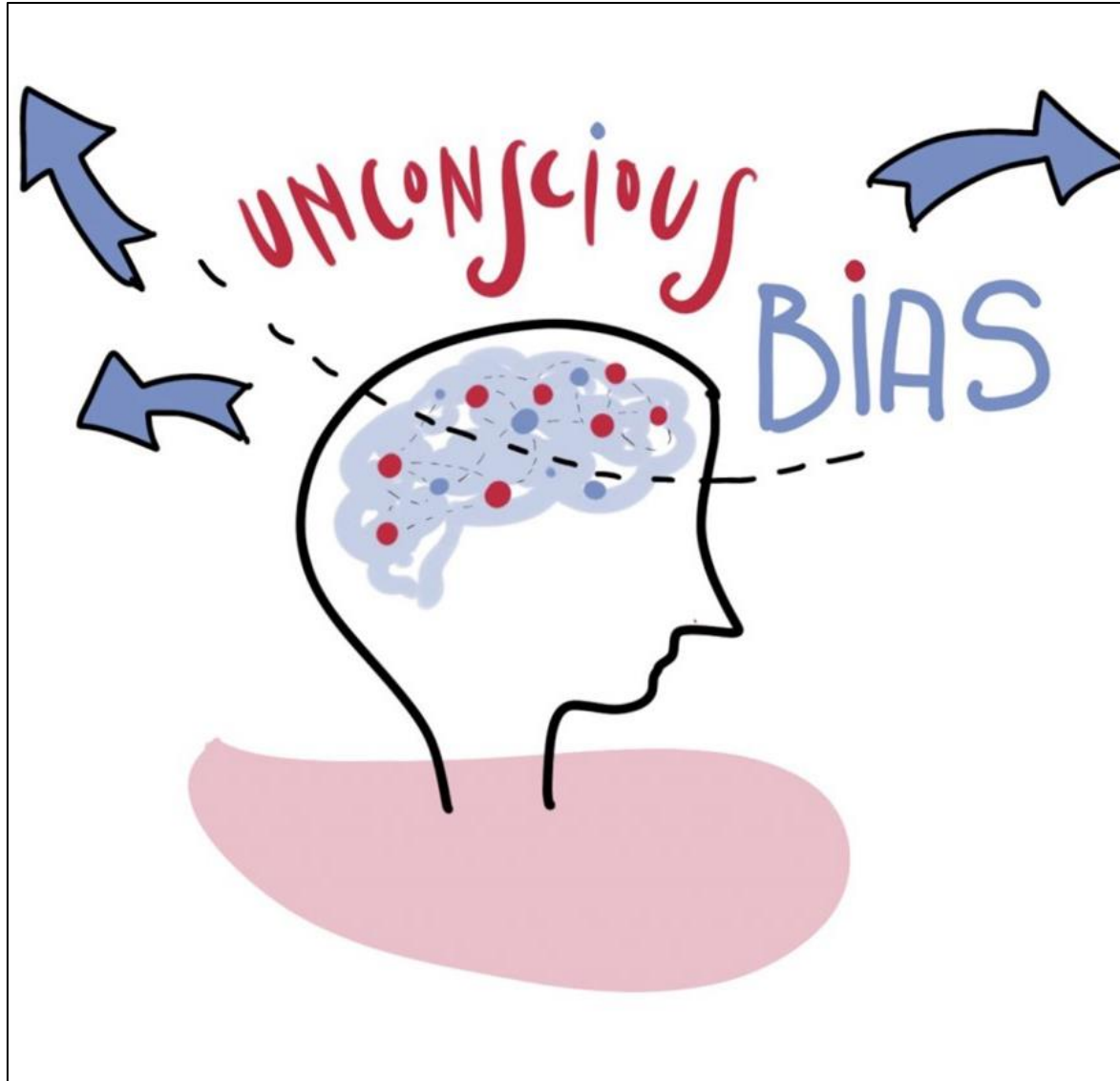
EXPLICIT PREJUDICE



IMPLICIT PREJUDICE



IMPLICIT PREJUDICE



Characteristics

- Automatic and often unconscious
- Activated rapidly
- Difficult to access through self-report
- Influenced by repeated social experiences
- Can operate without intentional prejudice

Key Idea: People may reject prejudice consciously while showing automatic bias.

Social judgments can operate outside conscious awareness, laying the foundation for modern implicit prejudice research (Greenwald & Banaji, 1995)



WHY EXPLORE THE IMPLICIT?

COMPARING EXPLICIT AND IMPLICIT PREJUDICE

	Explicit	Implicit
Awareness	Conscious	Often unconscious
Control	High	Lower
Measurement	Self-report	Indirect measures
Speed	Deliberate	Rapid
Social desirability effects	High	Lower

Theoretical Foundations of Implicit and Explicit Prejudice Interaction



Dissociation Model (Devine, 1989)



MODE Model (Fazio, 1990)



Dual Attitude Models (e.g., APE, Reflective-Impulsive Model)



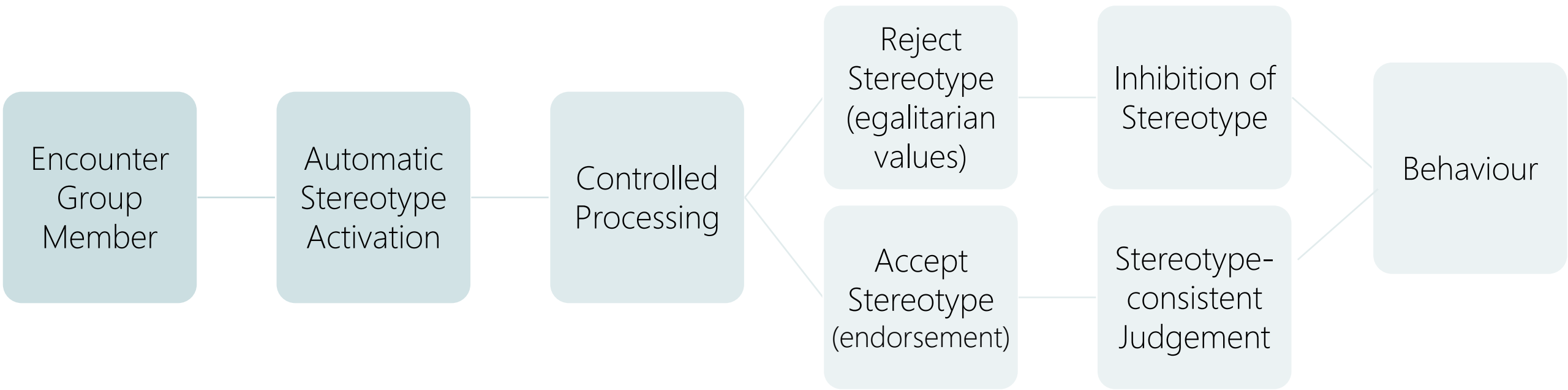
The two-dimensional model of Prejudice
(Son Hing, Chung-Yan, Hamilton and Zanna (2008))

DISSOCIATION MODEL (DEVINE, 1989)

Automatic Activation
of Stereotypes

Controlled Processing
of Responses

HOW THE DISSOCIATION MODEL WORKS



Key Idea

Both high- and low-prejudice individuals may activate stereotypes automatically.

TESTING THE THEORY: DEVINE (1989)

Participants

- High-prejudice individuals
- Low-prejudice individuals

Procedure

1. Complete prejudice scale
2. Exposed to stereotype-related primes
3. Read about an ambiguous target person
4. Evaluate the target

Prediction

Both groups would show evidence of stereotype activation.

FINDINGS OF DEVINE (1989)

1. Stereotypes were activated automatically
2. High and low prejudice participants showed similar stereotype activation
3. Personal beliefs influenced later responses

Conclusion

Knowledge of stereotypes is distinct from endorsement of stereotypes.

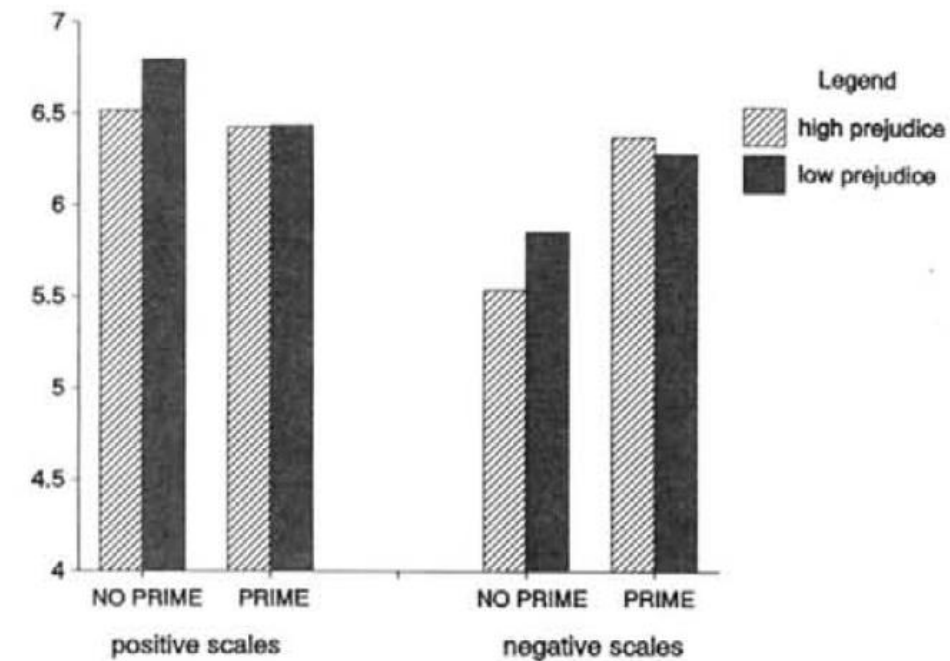


Figure 2. Priming effects on rating of target person.

STRENGTHS AND LIMITATIONS

Strengths

Highly influential theory
Explains implicit-explicit
attitude conflict
Supported by later
research

Underestimates
individual differences in
stereotype activation
Limited attention to
context and motivation
Automatic processes
may not be completely
uncontrollable

Limitations

MOTIVATION AND OPPORTUNITY AS DETERMINANTS OF BEHAVIOUR (MODE MODEL; FAZIO, 1990)

Motivation

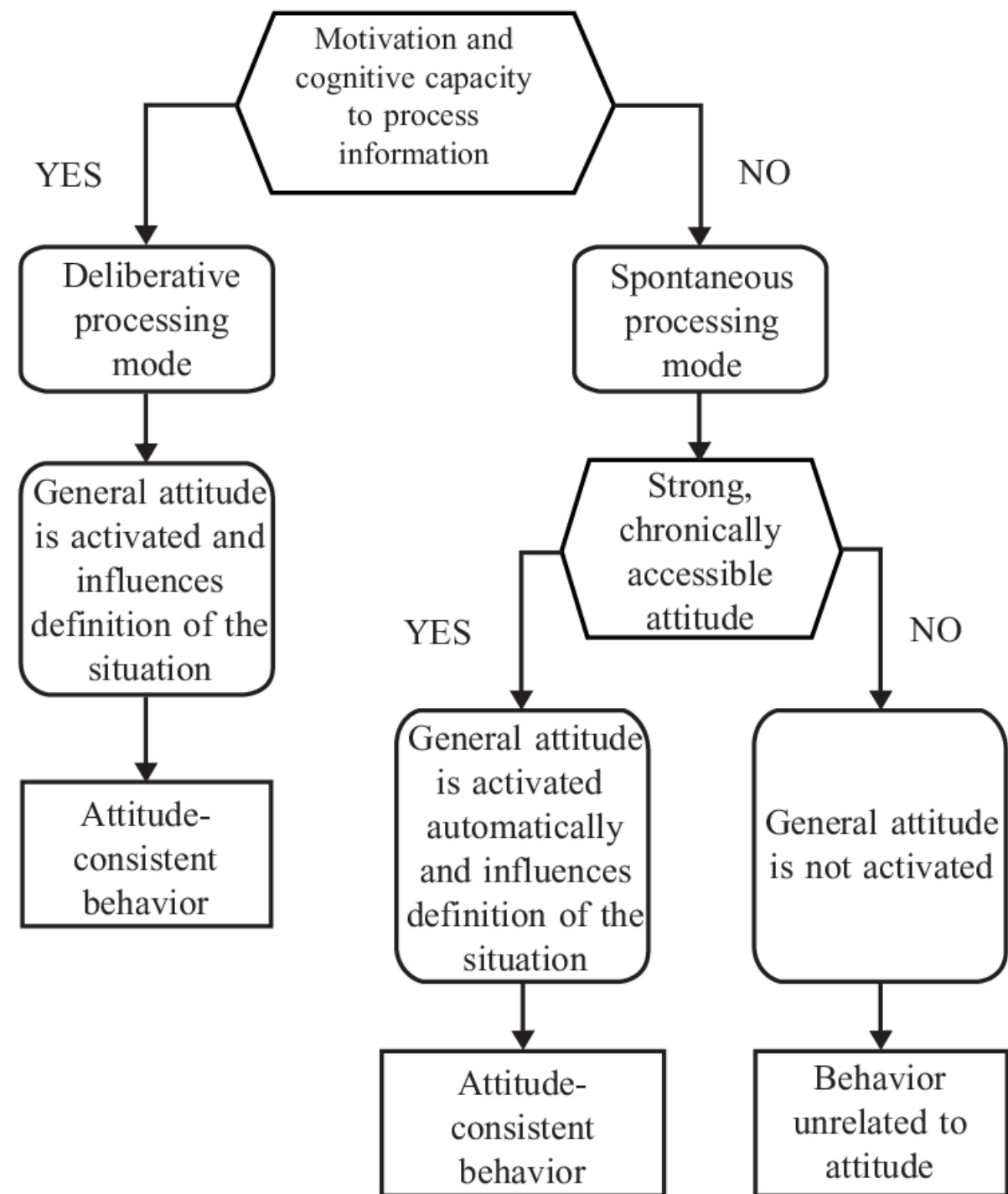
Do I want to control my
response?

Opportunity

Do I have the time and cognitive
resources to do so?

When motivation or opportunity is low, automatic attitudes are more likely to guide behaviour.

HOW THE MODE MODEL WORKS



WHEN ARE AUTOMATIC BIASES MOST LIKELY TO INFLUENCE BEHAVIOUR?

Motivation	Opportunity	Likely Behaviour
LOW	LOW	Automatic, biased behaviour
HIGH	LOW	Mixed or conflicted behaviour
LOW	HIGH	Automatic behaviour, possibly masked
HIGH	HIGH	Controlled, deliberate behaviour

STRENGTHS AND LIMITATIONS

Strengths

Explains when implicit biases influence behaviour

Integrates automatic and controlled processes

Supported by behavioural research (Fazio, Jackson, Dunton & Williams, 1995)

Difficult to measure motivation directly

Opportunity varies across situations

Boundaries between automatic and controlled processing may be unclear

Limitations

ASSOCIATIVE-PROPOSITIONAL EVALUATION MODEL (APE; GAWRONSKI & BODENHAUSEN, 2006)

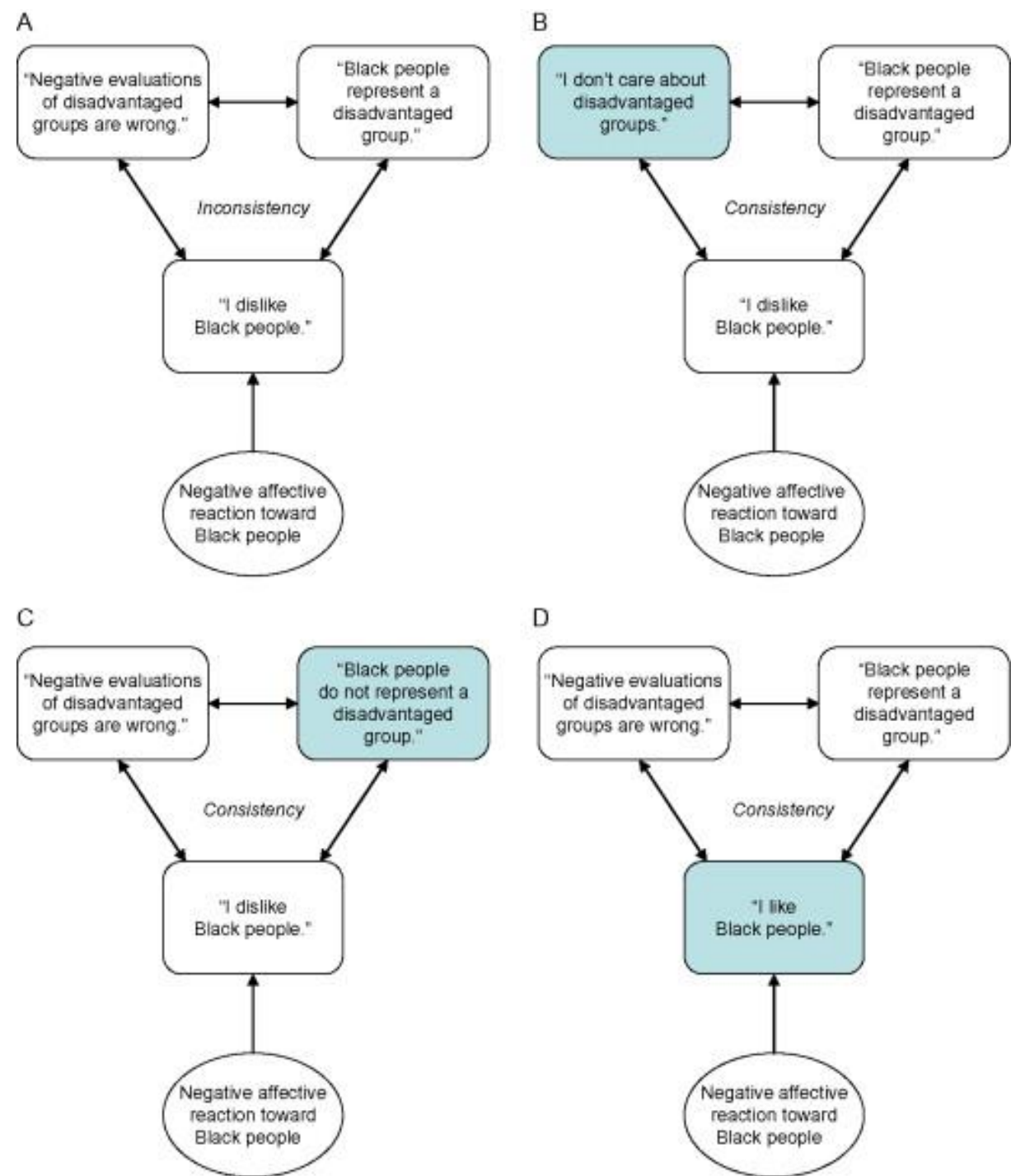
Associative Processes

- Automatic
- Fast
- Based on learned associations

Propositional Processes

- Deliberate
- Reflective
- Based on truth and reasoning

HOW THE APE MODEL WORKS



STRENGTHS AND LIMITATIONS

Strengths

Explains why implicit and explicit attitudes may differ

Integrates automatic and reflective processes

Explains attitude change and inconsistency

Difficult to distinguish associations and propositions empirically

Processes may interact more dynamically than proposed

Can be complex to test directly

Limitations

REFLECTIVE-IMPULSIVE MODEL (RIM; STRACK AND DEUTSCH, 2004)

Two interacting systems:

Impulsive System

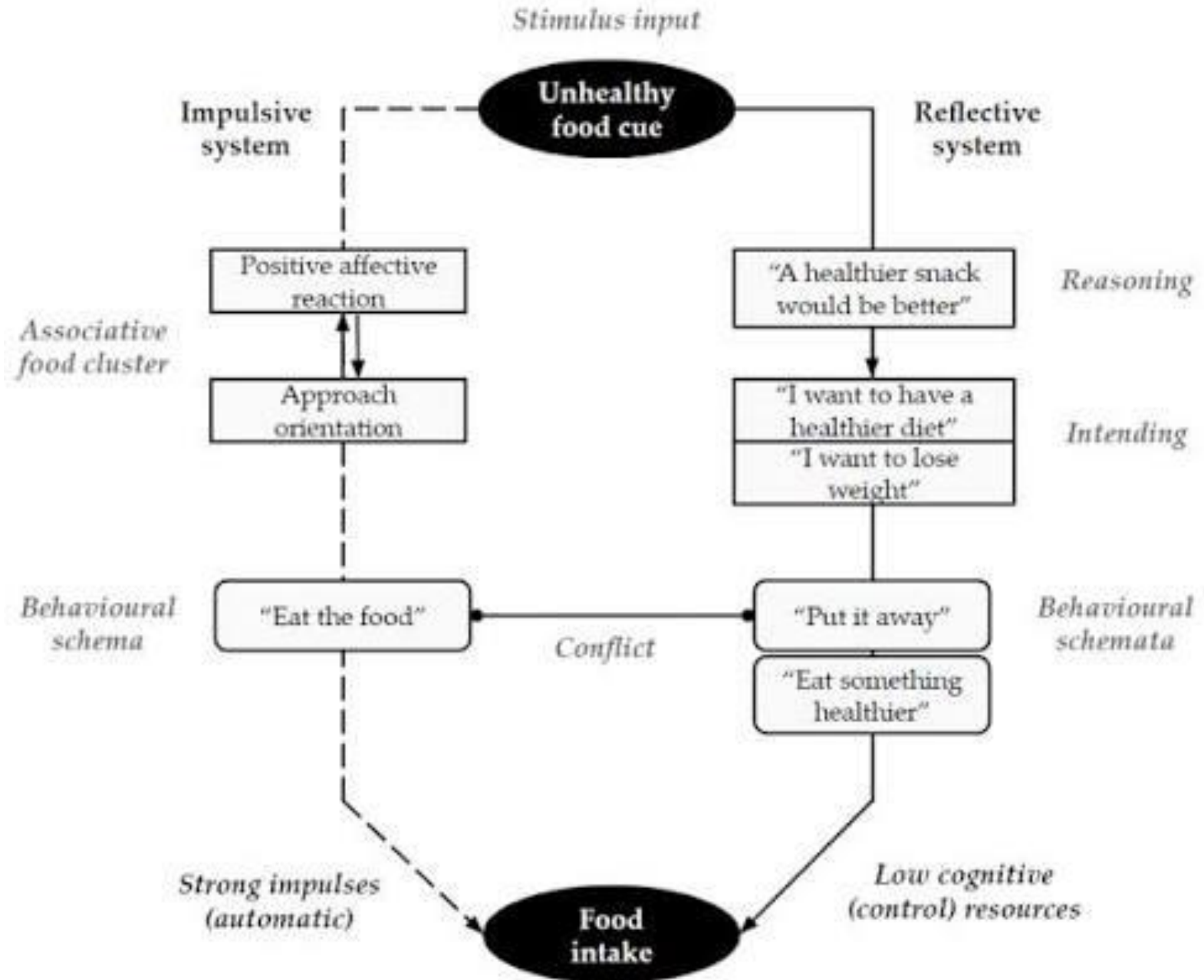
- Fast
- Automatic
- Associative
- Emotion driven

Reflective System

- Deliberate
- Reasoned
- Rule-based
- Conscious

Both systems simultaneously influence behaviour.

HOW THE REFLECTIVE- IMPULSIVE MODEL WORKS



APPLYING THE RIM TO PREJUDICE

Impulsive System

- Automatic stereotype activation
- Emotional reactions
- Learned associations

Reflective System

- Egalitarian values
- Social norms
- Conscious regulation

Outcome?

Depends on which process is strongest.

STRENGTHS AND LIMITATIONS

Strengths

Integrates cognition,
emotion and behaviour
Explains inconsistency in
prejudice
Applicable across many
social behaviours

Reflective and impulsive
systems may be difficult
to separate
Broad and complex
model
Some overlap with
other dual-process
theories

Limitations

COMPARING DUAL PROCESS MODELS

Model	Main Focus	Key Contribution
Devine (1989)	Automatic stereotype activation vs controlled beliefs	Explains why stereotypes can be activated without being endorsed
MODE (Fazio)	Motivation and opportunity	Explains when attitudes influence behaviour
APE (Gawronski & Bodenhausen)	Associative vs propositional processes	Explains why implicit and explicit attitudes can differ
RIM (Strack & Deutsch)	Reflective vs impulsive systems	Explains how both systems jointly shape behaviour

EMPIRICAL SUPPORT FOR DUAL-PROCESS MODELS

Dovidio, Kawakami, & Gaertner (2002)

- Explicit attitudes predicted verbal friendliness.
- Implicit attitudes predicted nonverbal behaviors (e.g., eye contact, body language).
- Participants with low explicit but high implicit bias showed inconsistent behaviour, often perceived as awkward or unfriendly.

Greenwald, McGhee, & Schwartz (1998)

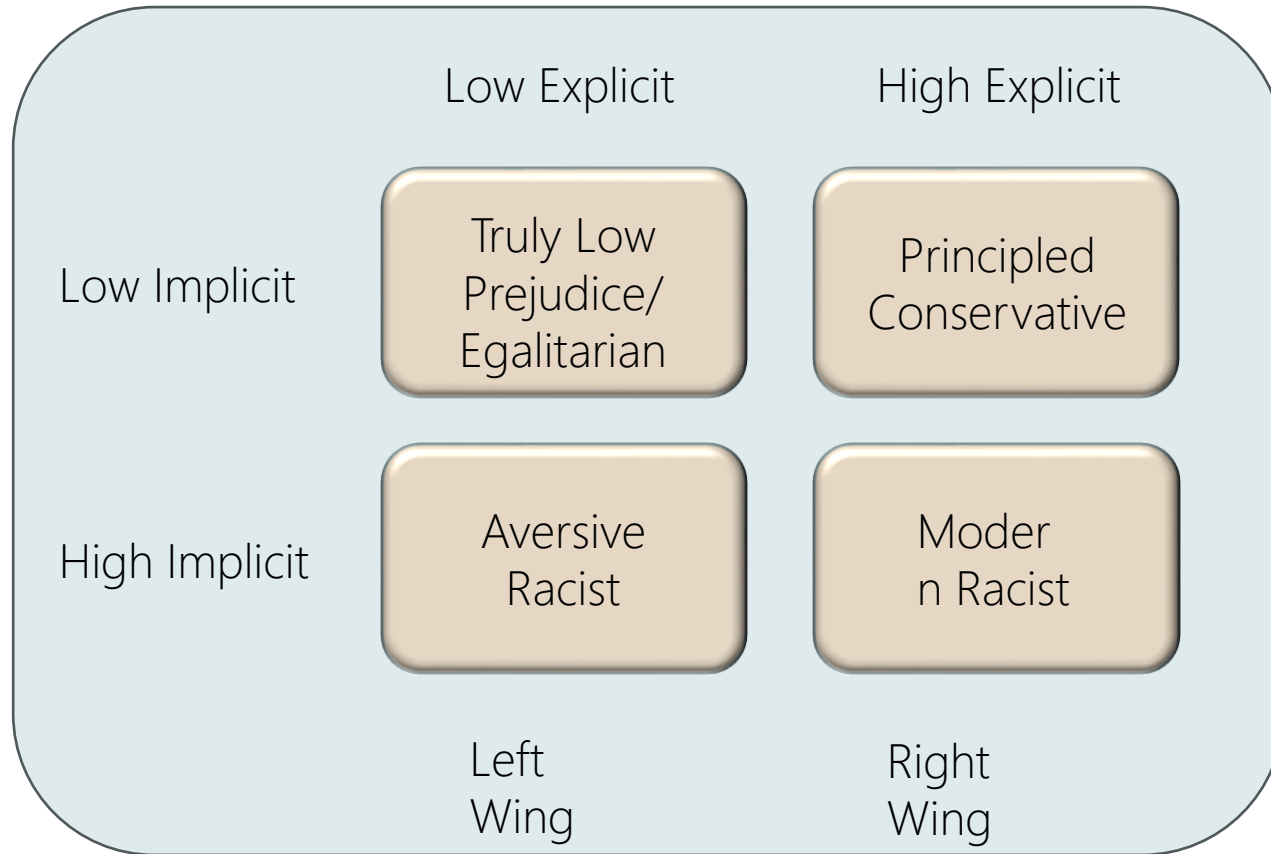
- Many participants showed implicit racial bias, even when their explicit attitudes were egalitarian.
- The IAT became a cornerstone for studying automatic stereotype activation.

Amodio & Devine (2006)

- Found distinct neural signatures for stereotype activation and evaluative bias.
- These processes influenced behavior in different ways, supporting the idea of multiple components in implicit prejudice.

Monteith (1993)

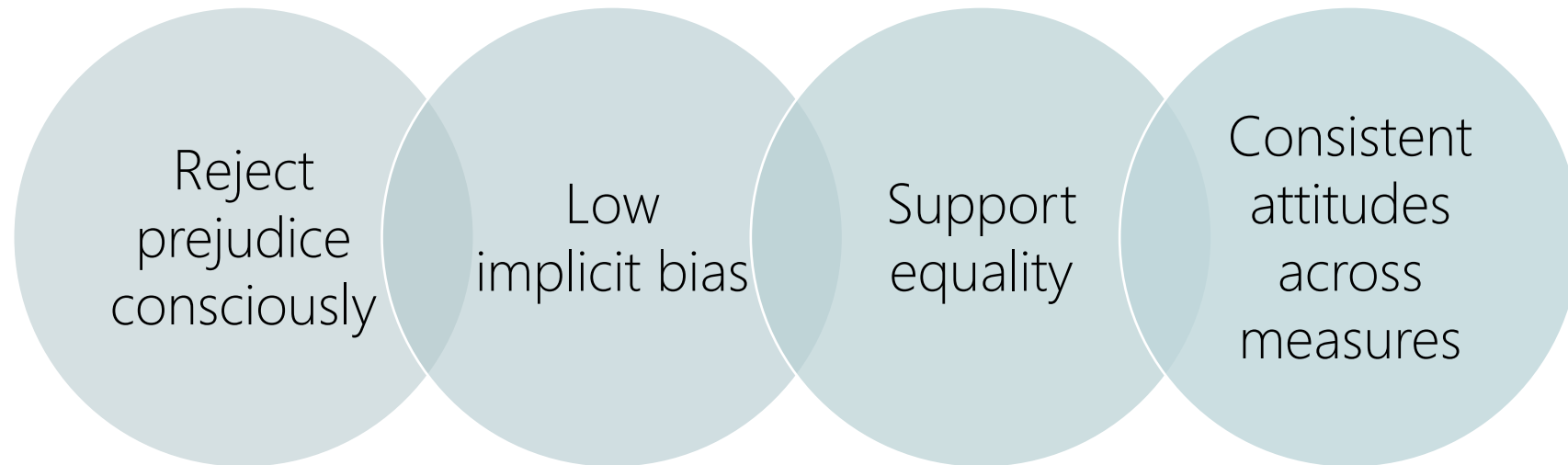
- Low-prejudice individuals experienced guilt and self-correction when they acted in stereotype-consistent ways.
- This supports the idea that controlled processes can override automatic activation.



THE TWO-
DIMENSIONAL MODEL
OF PREJUDICE
(SON HING, CHUNG-
YAN, HAMILTON AND
ZANNA, 2008)

LOW IMPLICIT & LOW EXPLICIT: TRULY LOW PREJUDICE/EGALITARIANS

Characteristics:

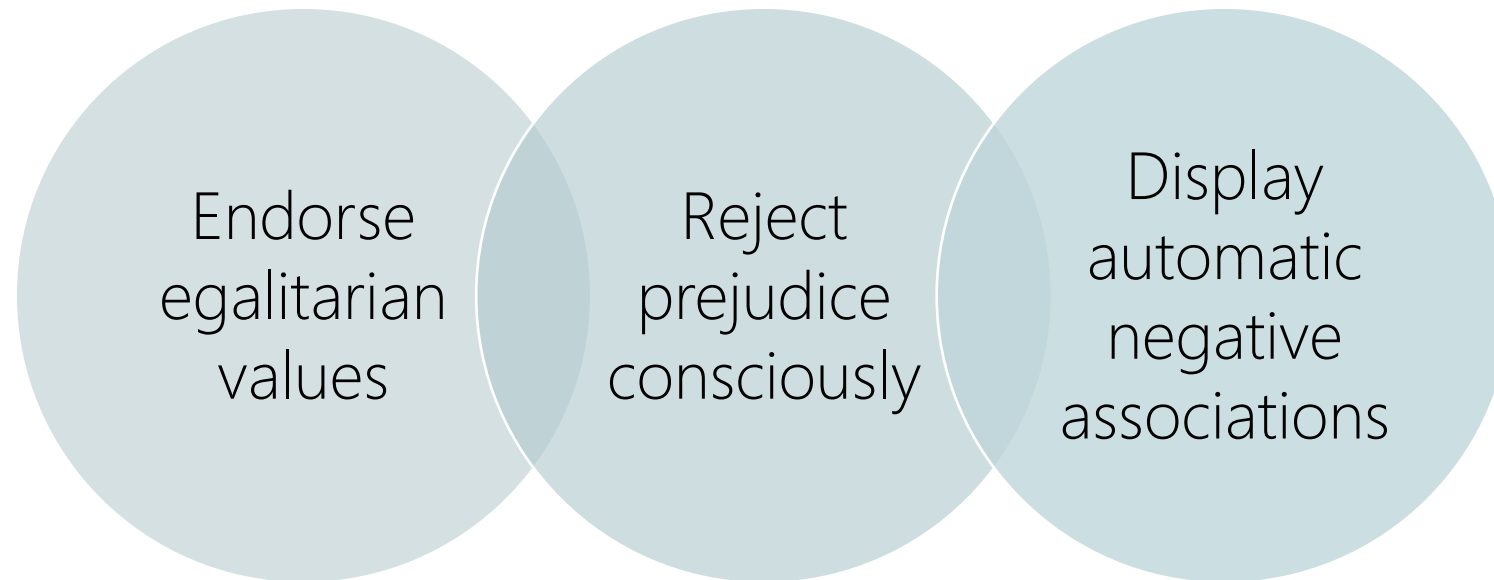


Behaviour:

Least likely to discriminate

HIGH IMPLICIT BUT LOW EXPLICIT PREJUDICE: AVERSIVE RACISTS

Characteristics:

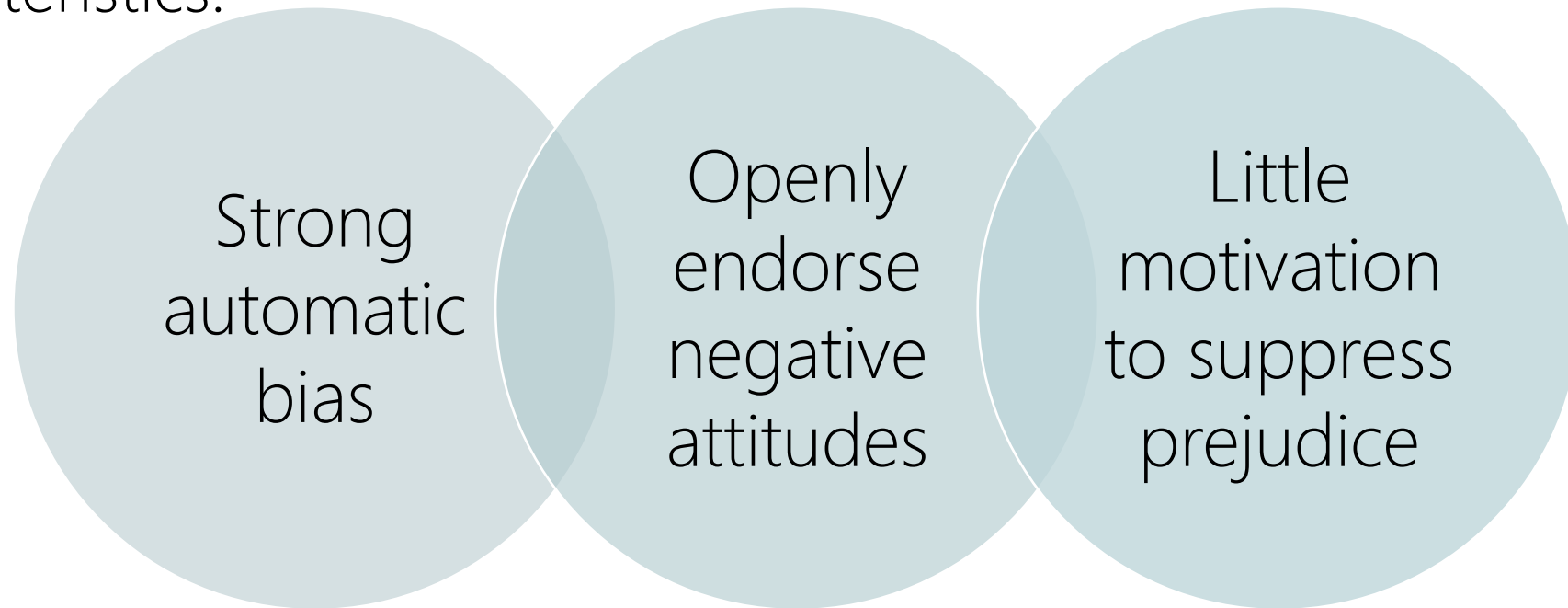


Behaviour:

- Bias most likely in ambiguous situations
- Behaviour often appears inconsistent

HIGH IMPLICIT & HIGH EXPLICIT: MODERN RACISTS

Characteristics:

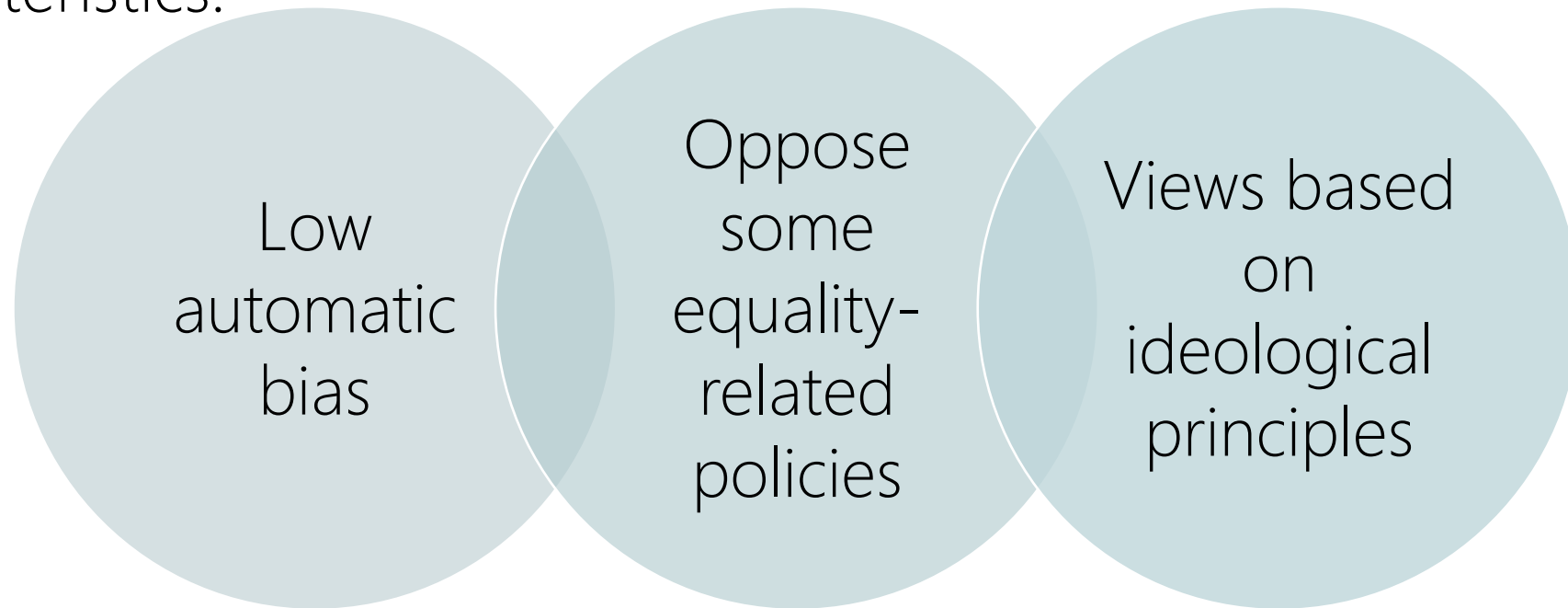


Behaviour:

Bias is more likely to be expressed directly

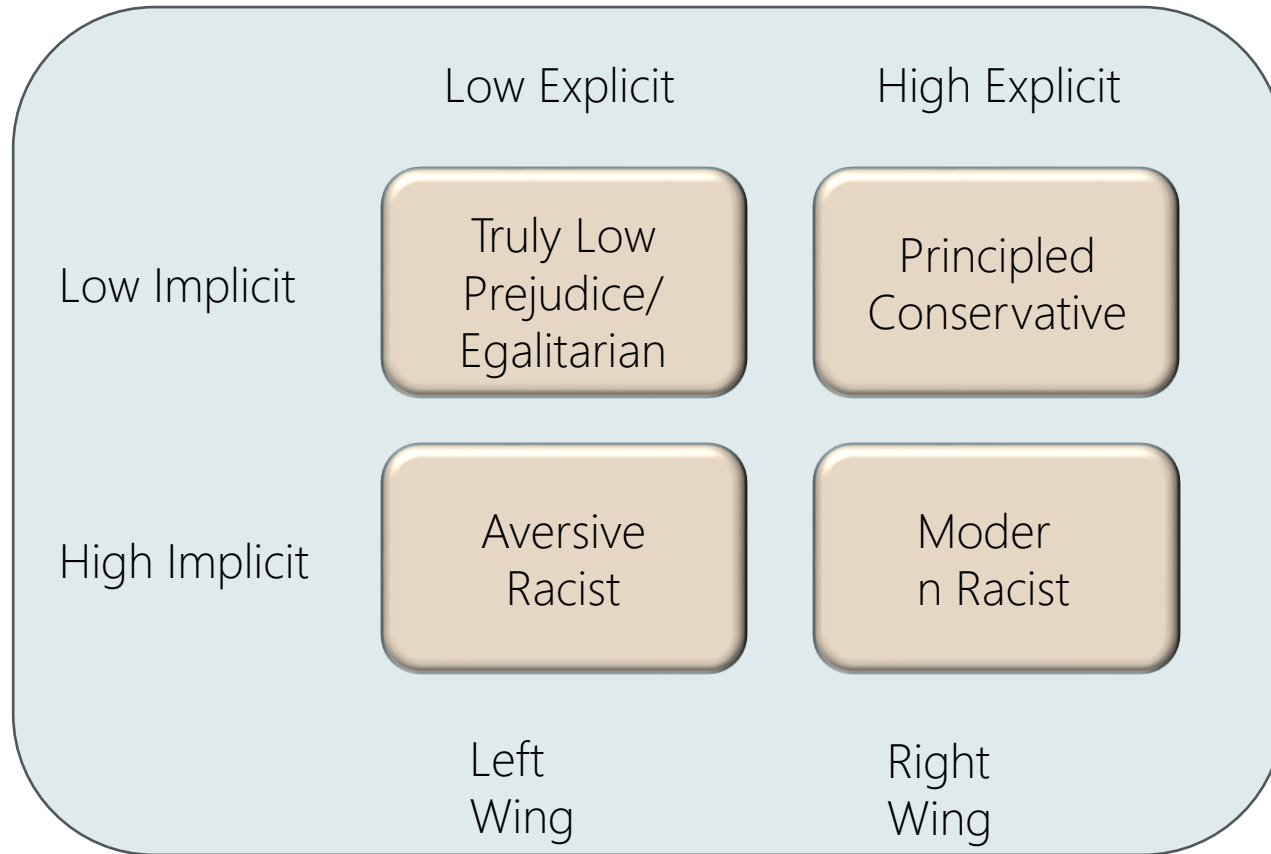
LOW IMPLICIT BUT HIGH EXPLICIT: PRINCIPLED CONSERVATIVES

Characteristics:



Debate

Prejudice or political ideology?



THE TWO-
DIMENSIONAL MODEL
OF PREJUDICE
SON HING, CHUNG-
YAN, HAMILTON AND
ZANNA (2008)

THE MEASUREMENT PROBLEM

Traditional measures rely on self-report.

But what if people:

- Cannot accurately access their attitudes?
- Do not want to reveal them?
- Are unaware of them?

Key Question

How can psychologists measure attitudes people may not consciously recognise?



MEASURES OF EXPLICIT PREJUDICE

ATTITUDE
TOWARD
BLACKS

(ATB; Brigham,
1993)

RACIAL
ATTITUDES
SCALE

(RAS; Sidanius,
Pratto, Martin, &
Stallworth, 1991)

SUBTLE AND
BLATANT
PREJUDICE
SCALE

(Pettigrew &
Meertens, 1995)

EXPLICIT
PREJUDICE
QUESTIONNAIRE

(Lepore & Brown,
1997)

MANY OTHERS

MEASURES OF IMPLICIT PREJUDICE

AFFECT
MISATTRIBUTION
PROCEDURE
(AMP)

Payne et al. (2005)

GO/NO-GO
ASSOCIATION
TASK (GNAT)

Nosek & Banaji
(2001)

EVALUATIVE
PRIMING TASK
SEMANTIC
PRIMING TASKS

Russell Fazio and
colleagues (1995)

NEUROSCIENTIFIC
MEASURES (E.G.,
FMRI, EEG)

MANY OTHERS

LOST LETTER TECHNIQUE

MILGRAM, MANN, & HARTER
(1965)



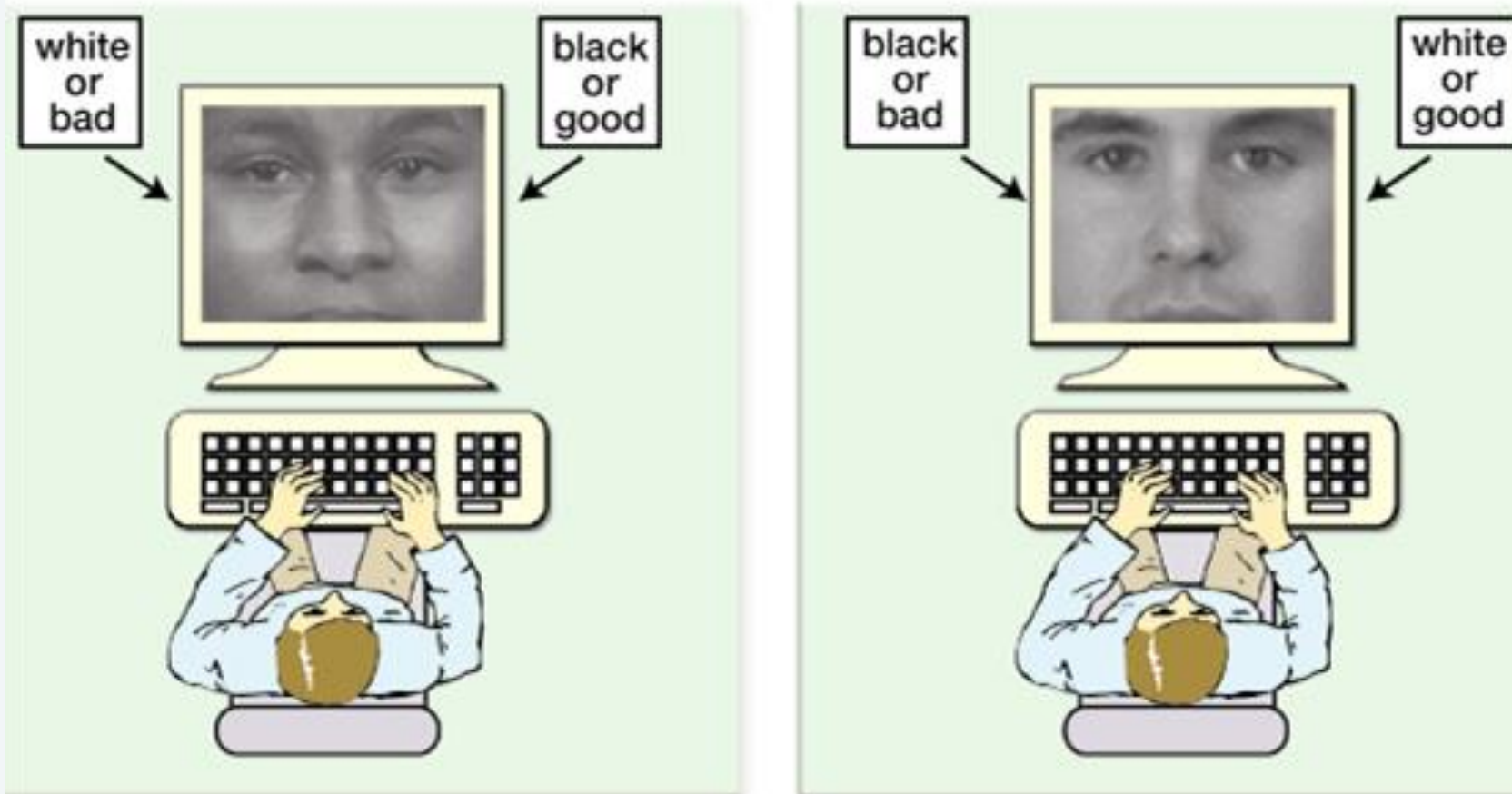
LOST EMAIL TECHNIQUE

STERN & FABER (1997)
TYKOCINSKI (2009)
AGERSTRÖM, CARLSSON, &
STRINIĆ (2021)



IMPLICIT ASSOCIATION TEST

GREENWALD, MCGHEE, AND SCHWARTZ (1998)



STRENGTHS AND LIMITATIONS

Strengths

Most widely used
implicit measure

Large research
evidence base

Predicts some forms
of behaviour

Modest predictive
validity

Sensitive to context

May reflect cultural
knowledge rather
than personal beliefs

Limitations

DOES IMPLICIT BIAS PREDICT BEHAVIOUR?

If implicit attitudes exist...

and can be measured...

Do they influence behaviour?

Competing Views



Implicit bias predicts discrimination



Implicit bias has limited behavioural effects

EVIDENCE SUPPORTING PREDICTIVE VALIDITY

Interpersonal interactions (Dovidio, Kawakami & Gaertner, 2002)

Explicit attitudes predict controlled, deliberate behaviours, whereas implicit attitudes predict more spontaneous behaviours.

Hiring Decisions (Bertrand & Mullainathan, 2004)

The findings are consistent with the argument that unconscious or automatic biases may contribute to employment discrimination.

Education (Van den Bergh et al., 2010)

The study suggests that hidden attitudes may influence educational expectations even when explicit prejudice is not reported.

Healthcare (Green et al., 2007)

This was one of the first studies to demonstrate:

- Dissociation between explicit and implicit attitudes.
- A direct relationship between implicit bias measures and healthcare decision-making.

EVIDENCE AGAINST STRONG PREDICTIVE EFFECTS

Criticisms

- Effects are often small
- Behaviour is influenced by many factors
- Explicit attitudes can outperform implicit measures
- Predictive validity varies across contexts

Alternative Explanations

- Social norms constrain behaviour
- Organisational procedures reduce bias
- Context shapes attitude expression
- Implicit measures may reflect cultural knowledge rather than personal beliefs

How much behaviour can implicit bias actually explain?

IMPLICIT ATTITUDES: LEGITIMATE BELIEFS OR CULTURAL KNOWLEDGE?

- Implicit Social Cognition Theory (Banaji & Greenwald (1995, 1998))
- Why Implicit Attitudes Are (Probably) Not Beliefs (Madva, 2015)
- Automaticity
- Implicit Attitudes and Beliefs (Kurdia, Manna, Charlesworth & Banaji, 2019)
- Entity Realism about Implicit Attitudes (Khalili & Kazemi, 2025)



WOULD JESSIE
JACKSON FAIL THE
IAT?



CRITIQUES OF IMPLICIT BIAS RESEARCH

Reliance on the IAT

Modest behavioural
prediction

Issues with reliability

Interpretation
remains debated

Public
understanding
often oversimplifies
findings

Nevertheless...

Implicit bias research transformed the study of prejudice.

KEY TAKE-HOME MESSAGES

- ① Implicit and explicit prejudice are distinct but related
- ① Dual-process models explain how hidden bias operates
- ① Implicit attitudes can be measured indirectly
- ① Behaviour reflects both automatic and controlled processes
- ① The role of implicit bias remains actively debated

The image features a decorative header and footer. The header at the top shows the back and side profiles of four people with different hair colors: blue, dark blue, light brown, and red. The footer at the bottom shows the front and side profiles of four people with different hair colors: red, dark blue, light brown, and red. The central area is a solid light gray background.

THANK YOU FOR LISTENING!