

Piliavin et al. (subway Samaritans)

AS

Read online at pastopic.com/cambridge/psychology-9990/notes/piliavin-et-al

Last checked 6 October 2026

© 2026 Pastopic. Free for personal study. Please share the link, not the file.

What you need to know

In 1969, Piliavin, Rodin and Piliavin staged a man collapsing on a moving New York subway train, again and again, and watched who came to his aid. Nobody on board knew it was a study. As with every AS core study, you should be able to:

1. **Explain the psychology being investigated:** **bystander apathy** and **diffusion of responsibility**, and what pushes people towards or away from helping when an emergency happens in front of them.
2. **Give the background:** the earlier cases and research that led to the study.
3. **State the aims** of the study.
4. **Describe the procedure and its methodology:** a **field experiment** using **observation**; the sample and how it was obtained; the victims, the models and the observers; the conditions (the independent variables); how each trial ran; what was recorded (the dependent variables); the intended experimental design; and the controls.
5. **Discuss the ethical issues:** deception, informed consent, protection from psychological harm, the right to withdraw, confidentiality and debriefing.
6. **Give the results:** the main quantitative findings (how often, how fast and by whom help was given) and the qualitative ones (what passengers said), and how they could be shown, for example in a bar chart.
7. **State the conclusions** the researchers drew, or that could be drawn.
8. **Evaluate every part of the study:** method, sample, data, controls, validity, reliability, generalisability and ethics.
9. **Relate it to the AS issues and debates:** application to everyday life, individual and situational explanations, nature versus nurture, and the use of children in research. No animals were used.
10. **Explain why it belongs to the social approach**, using the approach's two assumptions, and compare it with the other social core studies, Milgram and Perry et al.

Understand it

The psychology: why bystanders do or don't help

A **bystander** is anyone who sees an emergency happening but isn't directly involved in it. **Bystander apathy** describes what happens when such witnesses stand by: they see that someone is in trouble, yet they fail to step in. Examiners give one mark for the "bystander" idea (someone sees something happen) and one for the "apathy" idea (they hold back, or are less likely to step in).

Diffusion of responsibility is one explanation. When you are the only witness, the duty to act lands entirely on you. Add more witnesses and each person feels only a share of it, assuming that someone else will deal with it, so the bigger the crowd, the less likely any one person is to act. It is an **active** process (people actively shift the duty onto others), and it predicts that **more** witnesses means **less** help. "Fewer people means less help" is the reverse and doesn't describe it.

Piliavin et al. wanted to know whether these ideas, mostly tested in laboratories, held up in a real place where the emergency happened right in front of everyone.

Background

- **Kitty Genovese**: a real murder case; reports said **38** people saw her being killed and not one of them stepped in. It made psychologists ask why crowds fail to act.
- **Darley and Latané**: the chance that a listener would go to the aid of someone having an epileptic fit dropped if the listener believed others were also within earshot.
- **Latané and Rodin**: when the witnesses to a fall were strangers to one another, they helped less.
- **Bryan and Test**: watching another person give help made people readier to give help themselves (the effect of a **model**).
- **Blame and sympathy**: earlier writers had argued that if someone's trouble looks partly their own fault, onlookers feel less sorry for them and help less; other work found that onlookers sometimes think worse of victims (derogate them).
- **Gaps in the research**: nearly all of this work was done in **laboratories**, and participants often only **listened** to the emergency rather than watching it unfold.

The background is what was known **before** the study. In a background question, nothing about what Piliavin et al. did or found earns credit, and the cost-reward model doesn't count either, as it came **after** the study.

The aims

The main goal was to move bystander research **out of the laboratory**: to stage an emergency in an everyday place, where witnesses could **see** the victim for themselves rather than just hear him. Within that, they asked:

- Does it matter whether the victim seems to have **caused his own problem**? A man who appears drunk could be blamed for collapsing; a man who appears ill could not.
- Does the victim's **race** (black or white) change how often, and by whom, he is helped?

- Does seeing someone else help (a **model**) make others help, and does it matter **where** the model stands and **how soon** he acts?
- Does the **number of people** nearby change how often or how quickly help comes, as diffusion of responsibility predicts?

The method: a field experiment

It was an **experiment** because the researchers **manipulated** independent variables (the type of victim, his race and what the model did) and measured helping. It was a **field** experiment because it took place in an everyday setting, a subway carriage, and the participants were ordinary passengers going about their journey. It was **not** a natural experiment: the researchers created the conditions themselves. Data were gathered by **observation**: observers posing as ordinary passengers (**covert** observation) recorded what happened.

The sample

Feature	Detail
Size	About 4450 passengers
Where	New York City subway, on trains between Harlem and the Bronx , in either direction
When	Between 11 am and 3 pm
Who	Men and women; about 45% black and 55% white
Per carriage	On average 43 people in each carriage
Sampling technique	Opportunity : whoever happened to be riding in the carriage became a participant; nobody was asked or knew they were in a study

The team: victims, models and observers

Each trial used a **team of four students**: **two men** (one played the victim, one the model) and **two women** (the observers).

The victims (4 men, **3 white** and **1 black**, aged **26 to 35**). They all dressed the same way: an **Eisenhower jacket**, **old slacks** and **no tie**. Each victim played one of two roles:

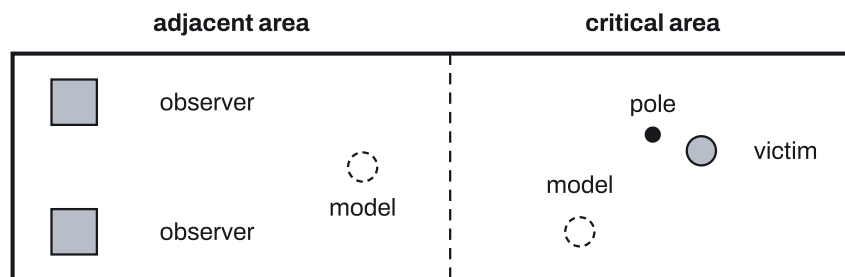
	"Drunk" victim	"Ill" (cane) victim
What he carried	A liquor bottle wrapped in a brown paper bag	A black cane
Smell and manner	Smelled of liquor	Appeared sober
Could he be blamed?	Yes: he seemed to have brought it on himself	No

The **cane** alone doesn't identify the ill victim for a mark; say he **appeared ill** and **carried a cane**. Both victims were men of the same age range: the June 2023 report says some answers wrongly made the ill victim old, the drunk victim young, or some victims female.

The models (4 men, all **white**, aged **24 to 29**, students from **Columbia University**). They wore **informal** clothes, and not all the same outfit. A model's job was to help the victim at a set point if nobody else had, so the researchers could see whether one person helping draws others in.

The observers (2 women). They sat in the **adjacent area**, outside the critical area, and recorded data as covertly as they could during the ride.

The carriage



Dashed circles: the model stood in one area or the other.

Schematic only: positions not to scale.

- The **critical area** is the part of the carriage around the pole where the victim stood and then collapsed.
- The **adjacent area** is the neighbouring part of the same carriage. Passengers there could also see the collapse.

The conditions (independent variables)

- **Type of victim:** drunk or ill (cane).
- **Race of victim:** black or white.
- **Model:** where he stood (**critical** or **adjacent** area) and when he helped (**early** or **late**), giving four model conditions, plus trials with **no model** at all.

The victim's condition was **not** chosen at random: every trial on a given day used the same type of victim, and teams were told to **alternate** the condition from one day to the next. (The June 2022 report says most answers wrongly wrote "random".)

What happened on each trial

1. The four students got on the train through **different doors**, and the carriage used was changed from trial to trial.
2. The two women sat down in the adjacent area. The victim and the model stayed **standing**.
3. The victim stood next to the **pole** in the middle of the critical area.
4. After the train passed the first station, about **70 seconds** into the ride, the victim **staggered forward and collapsed**.
5. He lay on the floor **face up** (supine) and stayed there until someone helped him.

6. The model, in the conditions with one, stood in his set area and stepped in at his set time if nobody had helped. In the **critical area – early** condition, he stood in the middle of the critical area and helped about **70 seconds** after the collapse.
7. If no passenger helped, the model helped the victim up.
8. At the next stop, the victim left the carriage.

What was recorded (dependent variables)

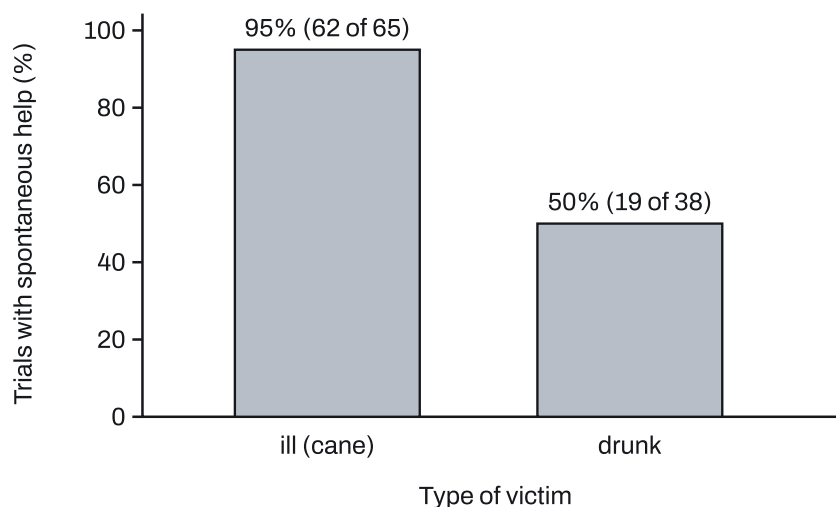
About	What the two observers recorded
Passengers in the critical area	Their race, sex and location , and how many there were
The helpers	The total number of people who came to help; the race, sex and location of every helper ; the latency (how long it took) before the first helper arrived
What people said	Comments passengers made

In the DV terms examiners use: **frequency of helping** (how often the victim was helped, and by how many people), **speed of helping** (latency in seconds), **who helped** (race, sex), **who moved out of the critical area**, and **comments** (qualitative data). **Spontaneous helping** means a passenger helped **before** the model stepped in.

The intended design

The design was meant to be **independent measures**: each passenger should take part in **only one** condition. In practice, a regular rider on this route could easily have been in the carriage for two trials (say, a drunk one on Monday and an ill one on Tuesday), and nobody was keeping track of who was who.

The results



How often help came

- The **ill** victim was helped spontaneously on **62 of 65** trials, about **95%**; the **drunk** victim on **19 of 38**, exactly **50%**. Help came far more often for the ill man.
- With **no model**, the black and white **ill** victims were both helped on **100%** of trials. A **white drunk** victim was helped more often than a **black drunk** victim, and the black drunk victim got the least help (**73%** of trials).
- On **60%** of trials more than one person helped.
- Most of the time, passengers helped **before** the model had a chance to.

How fast help came

- On average, the **ill** victim was helped **faster** than the drunk victim.
- Groups of **seven** helped **faster** than groups of **three**, and the more men there were in the critical area, the shorter the wait for help. The larger groups were not more **likely** to help, just **quicker** (March 2021 report).

Who helped

- About **90%** of the help came from **men**.
- **Same-race helping**: there was a slight, non-significant tendency for the ill victim to be helped by his own race, and a stronger one when the victim was drunk. For example, **68%** of spontaneous helpers of a white victim were white, which is higher than the proportion of white people in the carriage.

Leaving the critical area

- **34** people left the critical area, on **21** trials (**20%**).
- People were more likely to leave when the victim was **drunk** than when he was ill, and when no help had come within **70 seconds** of the collapse.

Comments (qualitative)

- More comments were made about the **drunk** victim than the ill one, and more when the victim was **white** than when he was black.
- Things said by people who didn't help included: "It's for men to help him", "I wish I could help him", "I am not strong enough", "I never saw this kind of thing before", "I don't know where to look" and "You feel so bad when you don't know what to do".

A **result** with a **meaningful comparison** says which condition got more or less of something than another ("the ill victim was helped faster, on average, than the drunk victim"). A one-sided statement ("the ill victim was helped quickly") scores only 1 mark.

The conclusions

- **Out in the real world, someone who collapses in plain sight is very likely to get help**, and usually quickly.
- **Diffusion of responsibility did not appear on the train**: bigger groups of passengers helped faster, not slower.

- **Blame matters:** when the man looked as though he had brought the collapse on himself, he was helped less, and less quickly, than when he seemed simply unwell.
- **In a physical emergency like this, it is mostly men who act.**
- **Race played a part mainly for the "drunk" man:** his helpers tended to share his race.
- **The cost-reward model:** whether a bystander helps depends on what they think helping would cost them (effort, embarrassment, danger) compared with what not helping would cost (guilt, blame). A drunk man may seem riskier to approach, so the cost of helping him feels higher.

Conclusions are **general claims** about people; numbers such as "62 of 65 trials" are results and score 0 when a conclusion is asked for.

Ethical issues

Guideline	What it means	In this study
Informed consent	Saying yes to a study knowing what it involves; nobody can agree to something they haven't been told about	Passengers had no idea a study was going on, so nobody could agree to it (broken)
Deception	Misleading participants on purpose	The collapse was staged; the "drunk" and "ill" men and the model were actors (broken)
Protection from psychological harm	Taking part must not leave anyone more upset, anxious or ashamed afterwards than they were beforehand	Seeing a man fall and lie on the floor could be upsetting; some who didn't help may have felt guilty afterwards (arguably broken). But nobody had to keep watching, and since help usually came quickly from someone, most onlookers had little to feel bad about
Right to withdraw	Anyone in a study can drop out at a moment of their choosing, and nothing bad follows	Nobody can step off a moving train, and they didn't know there was a study to leave (broken). On the other hand, the 34 people who walked out of the critical area had, in a sense, removed themselves from the situation
Confidentiality	Information collected must not reveal who any participant was	Nobody's name was ever taken; the only details recorded were things like sex, race and where people stood, and none of that could identify a passenger (kept)
Debriefing	Telling participants the truth afterwards	Passengers simply got off at their stops; no one explained the set-up to them, so most will have gone home believing they had seen a real collapse (not done)

Confidentiality and **privacy** are different guidelines: don't mix them up (March 2023 report).

Types of data

- **Quantitative:** how many trials had help; how many people helped; latency in seconds; the race and sex of helpers; how many left the critical area. Objective, easy to compare (for example, ill against drunk victims as percentages).
- **Qualitative:** the comments passengers made. These help explain **why** people did or didn't help: for instance, one non-helper said it was for men to help him.

Why it is from the social approach

The social approach rests on two ideas from the syllabus, put here in plain words:

- **Settings and groups:** the situation you are in and the groups around you shape what you do, what you think and what you feel (**behaviour**, **cognitions** and **emotions**).
- **Other people's presence:** others can change your behaviour, thoughts and feelings whether they are physically there, merely expected, or only imagined.

Piliavin et al. fits both. Passengers' **behaviour** changed with the social situation: they helped less, and more slowly, when the man seemed drunk; in model trials, the model's lead was what finally got certain passengers out of their seats; and the number of men nearby changed how fast help came. Their **emotions** show in comments such as "You feel so bad when you don't know what to do".

Issues and debates

- **Application to everyday life:** the findings suggest ways to get bystanders to help sooner: for example, training staff or volunteers to step forward first, since others may follow a helper, or campaigns reminding people that someone who has collapsed may be ill rather than drunk.
- **Individual and situational explanations:** an **individual** explanation puts the cause inside the person: their personality, attitudes or beliefs. A **situational** one puts it in the surroundings: the setting, the other people and what is happening. Situational: the victim's condition changed how often he was helped; the number of people nearby changed how fast help came. Individual: on the same train, some passengers rushed to help while others moved away from the critical area; one non-helper said it was for men to help.
- **Nature versus nurture:** mainly **nurture**: ideas about who deserves help, or that "it's for men to help", are learned from culture and upbringing.
- **The use of children:** no children were deliberately studied. Using them would raise problems: a child who saw a man collapse might be frightened, and could not give consent; and children may not be strong enough, or allowed, to help an adult.

Comparing with the other social studies

Point	Piliavin et al.	Milgram (obedience)	Perry et al. (personal space)
Method	Field experiment	Controlled laboratory study (no IV)	Experiments

Point	Piliavin et al.	Milgram (obedience)	Perry et al. (personal space)
Setting	New York subway train	University laboratory	Computer tasks: a distance scale, and choosing between pictures of rooms
Stooge	Students playing a fallen rider and a fellow passenger; everyone on board took them for real	A human stooge (the learner) met in person, then only heard	An animated, on-screen figure
Participants	About 4450 passengers, men and women, opportunity sample	40 men, volunteers	Participants given a hormone (oxytocin) or a placebo
Consent	None: unaware of any study	Agreed to a different study	Agreed to take part

Key facts and formulas

There is no formula list in the 9990 exam. "Learn it" means you must remember it.

Fact	Status
Piliavin sample: about 4450 New York subway passengers (Harlem–Bronx, 11 am–3 pm), men and women, about 45% black and 55% white, on average 43 per carriage, opportunity sample	Learn it
Piliavin victims: 4 men (3 white, 1 black), aged 26–35, Eisenhower jacket, old slacks, no tie; "drunk" smelled of liquor and held a bottle in a brown paper bag; "ill" appeared sober and carried a black cane	Learn it
Piliavin models: 4 white men, aged 24–29, informal clothes, Columbia University students	Learn it
Piliavin team per trial: 2 men (victim and model) and 2 women (observers)	Learn it
Piliavin timing: the victim collapsed about 70 seconds into the ride, after the first station; a model in an early condition helped about 70 seconds after the collapse	Learn it
Piliavin conditions: victim drunk or ill; black or white; model in the critical or adjacent area, early or late, or no model; victim condition alternated day by day	Learn it
Piliavin results: ill victim helped spontaneously on 62 of 65 trials, drunk on 19 of 38; about 90% of the help came from men; 60% of trials had more than one helper; 34 people left the critical area on 21 trials (20%)	Learn it
Percentage = $\frac{\text{part}}{\text{whole}} \times 100$, for example $\frac{19}{38} \times 100 = 50\%$	Learn it

How to do it

We have tagged 37 questions on this study, from 2021 to 2026 (Papers 1 and 2). 21 of them have parts of more than one type, so the counts add up to 60.

Question type	Questions
Results and conclusions	9
Describe the victims or the models	8
Research-methods ideas applied to Piliavin	8
Describe the procedure or what was recorded	6
One strength, weakness, validity or generalisability point	6
Background, aims and the psychology investigated	5
Social approach and issues and debates	5
Compare with another study or a new scenario	4
Evaluate in 10 marks: two strengths and two weaknesses	3
Ethics: define a guideline and apply it	3
Describe the sample	3

1. Results and conclusions (9 questions)

1. **Check which result is asked for:** drunk victims only, the no-model condition, race of helper, group size and speed, leaving the critical area, or comments. The reports say many answers gave a result from the wrong part, or a comparison the question didn't ask for.
2. **For 2 marks, make a meaningful comparison:** say what was higher or lower, faster or slower, **than what**. "The ill victim was helped more" is 1 mark; "the ill victim was helped more often than the drunk victim" is 2.
3. **When data are required, quote a number** (19 of 38; 100%; 73%; 90%; 34 people on 21 trials).

Variations you will meet:

- **The drunk victim before any model helped** (2 marks): helped on 19 of 38 trials; a white drunk victim was helped more than a black drunk victim; more comments were made when the drunk victim was white than when he was black; same-race helping was more common. Comparing drunk with ill earns only 1 mark here.
- **The no-model condition, with data** (3 marks): black and white ill victims helped equally, on 100% of trials (comparison + data); white drunk helped more than black drunk; the black drunk victim got the least help, 73%.
- **Race of helper when the victim was drunk** (2 marks): more same-race helping, whatever the victim's race. Not time taken.

- **Same-race helping in the cane condition** (2 marks): a slight tendency for helpers to be the victim's own race, but not significant. Don't compare across victim types.
- **Number of males and time to help** (2 marks): the more males in the critical area, the faster help came; groups of 7 were faster than groups of 3. Not frequency, and not male versus female.
- **Time to help** (2 marks): the ill victim was helped faster on average than the drunk victim; groups of 7 were faster than groups of 3.
- **Leaving the critical area** (2 marks): more people left when the victim was drunk than when he was ill; more left when nobody had helped within 70 seconds; 34 people on 21 trials (20%) is a result without a comparison.
- **Comments** (1 mark each): more comments about the drunk victim than the ill one, and about white victims than black; non-helpers' comments, such as "I am not strong enough". Only real comments from the study count, not ones a passenger might have said.
- **A finding against diffusion of responsibility** (2 marks): larger groups helped faster than smaller ones; most passengers helped before the model; on 60% of trials more than one person helped.
- **A conclusion** (2 marks, when one is printed in the question): give a **different** one from "The conclusions" above. A full general claim, such as "bystanders are more willing to aid someone who looks unwell than someone who looks as if he brought the trouble on himself", earns 2; "ill victims got more help" earns 1. Repeating the printed one, or giving a number, earns 0.

2. Describe the victims or the models (8 questions)

The mark schemes call these lists **definitive**: only the facts in them score, one mark each.

- **The victims** (up to 4 marks): male; aged 26–35; three white and one black; dressed identically (Eisenhower jacket, old slacks, no tie: two of these for 1 mark); one type acted drunk, smelled of liquor and held a bottle in a brown paper bag; the other acted ill, appeared sober and carried a black cane. Write "**acted** drunk" or "**acted** ill", not "was drunk".
- **The drunk victim** (2 marks): at least one point must be the liquor smell or the bottle in a brown bag; the other can be his age, sex, race or clothes.
- **Three differences between the two types** (3 marks): drunk versus ill; a bottle in a bag versus a black cane; smelled of liquor versus sober.
- **Why the similarities mattered** (3 marks): it was a control (standardisation) that removed confounding variables (1); then the only thing changing between trials was the IV, drunk versus ill (1); plus a Piliavin-specific reason, such as matching clothes meaning neither man stood out from the other riders, or the model conditions running the same way whichever victim was used (1).
- **The models** (1 mark each): white; male; aged 24–29; informal clothes; Columbia University students; there were four. "Adults" earns nothing. Don't give the victims' features.
- **Differences between the models** (1 mark each): how old they were; what they wore; where they stood (critical or adjacent area); when they helped (early or late). The June 2022 report says many answers gave victim differences instead.

3. Research-methods ideas applied to Piliavin (8 questions)

Each is a definition (1) plus a Piliavin link (1 or 2). Define the term without using the term itself.

- **Independent measures** (1 mark): each condition of the IV is done by its own separate group of people.
Why it may not have been (1): a commuter could have been in the carriage on two trials with different victim or model conditions, and nobody tracked who had already been tested.
- **Operationalisation** (3 marks): turning a variable into something concrete that can be set up or counted the same way every time (1). IV: the "drunk" level was a man smelling of liquor with a bottle in a bag, the "ill" level a sober man with a cane (1). DV: counted as **spontaneous** help, a passenger reaching the victim **ahead of the model** (1). Operationalise **both** levels of the IV.
- **Participant variables** (2 marks): characteristics that vary from one participant to another, such as mood or personality, which can change their responses and so blur the effect of the IV (1); for example, a tired, irritable passenger might ignore any victim, while a naturally kind one would help either (1). "Time of day" is a situational variable, not a participant one.
- **A bar chart** (4 marks): y-axis labelled with what was counted, such as "number of victims helped"; x-axis "type of victim" with separate bars for "ill (cane)" and "drunk"; bars plotted accurately.
- **The aim of a study** (1 mark): the purpose behind it, the question the researcher wants answered. One Piliavin aim (1): to test whether the victim's apparent responsibility (drunk or ill) affects helping; whether race affects helping; whether models encourage helping; to test diffusion of responsibility.
- **Problems hearing comments** (1 mark): if passengers spotted someone listening in, they might go quiet; an observer trying to stay unnoticed on a loud, moving train could miss words; nothing was taped, so comments relied on memory; remarks in a language the observer didn't speak would be lost.
- **Controls** (see the victims' identical clothes, the fixed 70 seconds, the alternating days, the same pole) and **sampling**: opportunity sampling, because the participants were whoever happened to be on the train.

4. Describe the procedure or what was recorded (6 questions)

One mark per correct fact, so give separate, specific points inside the start and end points the question gives.

- **From boarding until the collapse** (5 marks): two men and two women; boarded through different doors; the carriage varied; the women sat in the adjacent area; the victim and model stood; the victim stood by the pole in the critical area; the model stood in the critical or adjacent area depending on condition; observers recorded covertly; the victim waited about 70 seconds before collapsing. One mark at most for describing the victim.
- **What the victim did** (5 marks): stood by the pole; in the critical area; after the first station (about 70 seconds); staggered forward and collapsed; stayed lying face up until helped; if nobody helped, the model helped him up; got off at the next station. Not the differences between victims.
- **The model in the critical-area, early condition** (2 marks): stood in the middle of the critical area; waited about 70 seconds; then helped the victim.
- **What the observers recorded** (up to 4 marks, or 2 marks for two measures): race, sex and location of passengers; the number in the critical area; the number of helpers; race, sex and location of every helper; latency of the first helper; comments. The June 2023 report says some answers gave the victim's features

instead.

- **How the victim condition was chosen** (1 mark): alternated by day. Not random.

5. One strength, weakness, validity or generalisability point (6 questions)

"Identify" needs just the name (1 mark); "outline" or "explain" adds a Piliavin example (1). The reports say a strength with no example from the study can't get full marks.

- **Strengths: ecological validity or mundane realism** (people really do collapse on trains, and passengers didn't know it was staged, so their behaviour was natural); **standardisation**, so it can be **replicated** (victims and models followed set instructions, dressed alike, collapsed at the same point); a **large sample** of about 4450, men and women, black and white.
- **Weaknesses: ethics** (deception, no consent, no debrief); **lack of control** (passengers' positions, the number in the carriage and their moods couldn't be controlled); **generalisability** (one city, one subway line); **difficult to replicate exactly** (each trial had different passengers); **observer bias**. "Generalisability" or "ethical issues" alone earns 0; say "**lacks** generalisability" or "broke the guideline of deception".
- **Two friends disagree about validity** (4 marks: name the type of validity, then evidence and explanation).
Has validity: because every victim wore the same clothes, a difference in helping between drunk and ill trials can't be put down to what he wore (internal validity); and a real train ride gives ecological validity.
Lacks validity: passengers differed in ways nobody could control (some may have dealt with emergencies before), which muddies whether victim type caused the helping; all the models were white; and four different men played the victims, who may not have acted identically. Reliability, quantitative data and "field experiment" don't count as validity points.
- **Two friends disagree about generalisability** (4 marks). **Can generalise:** with roughly 4450 riders, male and female, black and white, the sample is far bigger and more varied than a typical laboratory group, so its pattern of helping is more likely to hold for people in general. **Can't:** everyone was riding the same stretch of one city's subway, so the findings say little about, say, villagers or people from cultures with other norms, and a commuter route may carry the same riders again and again; the victims were all men, and only black or white. Say **who** the results can or can't be generalised to.

6. Background, aims and the psychology investigated (5 questions)

- **Background** (5 marks, one per point): see "Background". At most 2 marks for Kitty Genovese (one for naming her, one for a detail). Definitions of bystander apathy or diffusion of responsibility can earn credit. Nothing about what Piliavin et al. did or found, and no cost-reward model. The March 2026 report says many answers stuck to Kitty Genovese alone.
- **Define bystander apathy** (2 marks): the bystander part (someone is present and can see trouble happening) (1) and the apathy part (they fail to intervene, or are less likely to help) (1).
- **Define diffusion of responsibility** (2 marks): each extra onlooker lowers the chance that any one of them steps in (1), since everyone expects that somebody else will act (1). "Responsibility is shared" alone is brief (1).
- **An aim** (2 marks for a full one): see "The aims". "To investigate bystander behaviour" is brief (1). If the question rules out race, don't use race; gender was not a main aim.

7. Social approach and issues and debates (5 questions)

- **An assumption** (2 marks): a full assumption names **two** of behaviour, cognitions and emotions **and** what influences them (social settings and groups, or the presence of others). "Behaviour is affected by other people" is brief (1).
- **Why the study supports it, or why it is from the social approach** (2 marks): a Piliavin result (1) clearly linked to the assumption (1). For example: the drunk-looking man, who could be blamed for his own fall, drew help on fewer trials, so the social situation changed what people did; or another person's presence mattered, as some only helped once the model had. Linking to an **aim** rather than a finding earns 1 at most.
- **Two assumptions with a core-study example each** (6 marks): 2 for each full assumption plus 1 for each linked example. Piliavin: in model trials, certain passengers stayed put until the model went to the victim, and only then moved to help.
- **Individual and situational** (4 marks): define each without repeating the word (1 + 1), and give a Piliavin example of each (1 + 1); see "Issues and debates". "Situational explanations are about the situation" scores 0.
- **Why the study supports a situational explanation** (2 marks): first say what this kind of explanation claims, that the causes of an action lie in a person's circumstances rather than in their character (1). Then give a Piliavin case: the very same kind of passengers acted differently depending on whether the man on the floor smelled of liquor or carried a cane; or the size of the crowd altered the speed of help (1).

8. Compare with another study or a new scenario (4 questions)

- **Why passengers didn't help in a new incident** (4 marks, one per point): use Piliavin knowledge: perhaps the person looked drunk, and onlookers judged him undeserving or worried he'd turn violent; his race; the sex of the people nearby (if most were women, they may have expected men to help, as "it's for men to help"); diffusion of responsibility if many were present. Explain why they **didn't** help, not why people do. Don't argue racial prejudice: Piliavin found little evidence of it.
- **How a new study is similar to or different from Piliavin** (1 mark each): a similarity must be specific (both about helping; both use stooges; two types of victim; participants are passengers on public transport and don't know they're in a study), not "both are field experiments". A difference must compare explicitly ("Piliavin used a model; this study doesn't").
- **Similarity and difference with Milgram or Perry** (4 marks each, levels): Level 4 explains the point using **both** studies. For example, a difference in the stooge: Piliavin's stooges were students posing as a fallen rider and a fellow passenger, and nobody on the train realised; Perry's was an animated figure on a screen. If one point must be about the **sample**, compare the actual samples (who they were, how many, where from), using facts you know for each study, not just the sampling technique.
- **How Piliavin differs from laboratory research** (2 marks each): name the difference, then say explicitly how the other research differs: Piliavin was a field experiment in a subway car, **unlike** the laboratory setting; Piliavin found more same-race helping, **unlike** the cross-race helping in the laboratory. The June 2021 report says the examiner can't make the link for you.

9. Evaluate in 10 marks: two strengths and two weaknesses (3 questions)

The named issue has been qualitative data, quantitative data and ethics.

1. **Plan four points of equal depth**, one on the **named issue**. Without it, the answer is capped at 6 marks.
2. For each point: name it, say whether it's a strength or weakness of **this** study, give **at least two** Piliavin details, and explain **why** it matters.
3. **No description** of the study for its own sake. Points without Piliavin context stay at the bottom level; a prepared "GRAVE" essay that misses the named issue is capped.
4. Strengths of qualitative data must say **why** they help: they explain helping or not helping (two examples make it detailed).

Point	Strength or weakness	Evidence from the study
Ecological validity	Strength	A real subway journey; an event that does happen; passengers didn't know it was staged, so no demand characteristics
Standardisation (reliability)	Strength	Victims dressed alike; always by the pole; collapsed after about 70 s; stayed down until helped; models at fixed times and places
Quantitative data	Strength	Frequency and latency of help, race and sex of helpers; ill versus drunk can be compared objectively (about 95% versus 50% of trials)
Qualitative data	Strength	Comments explain why some didn't help ("It's for men to help him", "I am not strong enough")
Sample size and variety	Strength	About 4450 passengers, both sexes, black and white
Ethics	Weakness	No consent; deception; a distressing sight; couldn't leave a moving train; no debrief
Control of variables	Weakness	Couldn't control who was in the carriage, where they stood, or whether they'd seen a previous trial
Generalisability	Weakness	One city, one route, in the middle of the day; victims all men, only black or white
Quantitative data	Weakness	Counts and times show what happened but not why

10. Ethics: define a guideline and apply it (3 questions)

Each guideline gets a definition (1) and a Piliavin example (1). The March 2023 report says definitions that just repeat the name ("right to withdraw means you can withdraw") earn nothing.

- **Right to withdraw:** a person may quit partway through, with no penalty, and ask for their results to be taken out (1); passengers couldn't get off a train between stations, and had no idea there was a study to quit (1). Or argue the other way: those who walked out of the critical area had partly removed themselves.

- **Deception:** giving participants a false picture on purpose (1); passengers thought they were watching a genuine emergency, but the fallen man and the model were both students playing parts (1).
- **Informed consent, protection from harm, confidentiality:** see the table under "Ethical issues". For confidentiality, it was **not** broken.
- **One ethical weakness** (2 marks): name the issue (1) and give the Piliavin example (1). A methodological weakness scores 0.

11. Describe the sample (3 questions)

One mark per fact: opportunity sample; about 4450; New York subway passengers; 45% black and 55% white; men and women; never asked or told; between Harlem and the Bronx; 11 am to 3 pm; on average 43 per carriage. The reports say answers often described the victims, the models or the procedure instead. "About 4450" is the number to state when asked how many took part.

Worked examples

Example 1

From the study by Piliavin et al. (subway Samaritans), explain what was meant by "spontaneous helping" and why the model's timing was needed to measure it. [3]

"Spontaneous helping" meant a passenger came to the victim's aid **before** the model stepped in (1). In the model conditions, the model held back for a set time (about 70 seconds after the collapse in the early conditions), which gave passengers a window in which any help came from them alone (1). Without that delay, the researchers couldn't tell whether a passenger had acted on their own or was only copying the model, so the DV would have mixed two different things (1).

Example 2

A railway company wants more passengers to help when someone faints on a crowded platform.

Using your knowledge of the study by Piliavin et al. (subway Samaritans), suggest two ethical ways the company could increase helping. [4]

1. The company wants someone to step forward quickly whenever a passenger faints (**what**, 1). It could train uniformed staff or volunteers to be the first to go over, since in Piliavin's model trials a few passengers stayed put until the model went over and only then moved, so one person acting can draw in others (**how**, 1).
2. The company wants people to help even when they aren't sure why the person fell (**what**, 1). Posters could explain that someone who collapses may be ill, not drunk, because Piliavin found a man who seemed drunk got help on far fewer trials than one who seemed ill (**how**, 1).

Both are ethical: no one is deceived, put at risk or punished.

Example 3

Explain why the study by Piliavin et al. (subway Samaritans) is classed as a field experiment and not a naturalistic observation. [3]

The researchers **manipulated** independent variables themselves: they decided whether the victim acted drunk or ill, whether he was black or white, and where and when the model helped (1). This happened in an ordinary, everyday setting, a moving New York subway carriage, with passengers who didn't know a study was going on, which makes it a **field** experiment (1). In a naturalistic observation, researchers only watch behaviour that happens by itself and change nothing; here the collapse was staged on purpose, so it was not simply observed (1).

Example 4

Evaluate Piliavin et al.'s subway study, giving two strengths and two weaknesses. Make reliability one of your four points. [10]

Plan: strengths are standardisation (reliability, the named issue) and ecological validity; weaknesses are ethics and generalisability. Each paragraph names the point, gives two or more details, and says why it matters.

Strength 1: reliability through standardisation. Every trial followed the same script. The victim always stood by the pole in the critical area, waited until the train had passed the first station (about 70 seconds), then staggered and fell and stayed on his back until someone helped. All the victims wore the same Eisenhower jacket, old slacks and no tie, and in the early model conditions the model waited about 70 seconds before stepping in. Because these steps were fixed, another team could run the identical procedure on a different line and compare results; if the pattern held, we could trust it as consistent, which is what reliability means.

Strength 2: ecological validity. The study took place on real trains between Harlem and the Bronx in the middle of the day, and someone collapsing on a train is something passengers might really face. Since nobody knew the collapse was staged, there were no demand characteristics: passengers weren't trying to behave as they thought the researchers wanted. Their helping, or their moving away, is therefore likely to show how people really behave in such an emergency.

Weakness 1: ethics. The collapse was a deception from start to finish: the victim and the model were students playing parts. Since nobody on board knew research was happening, informed consent was impossible. Some passengers may have found the scene distressing, and anyone who stayed in their seat could have carried guilt home with them. Because people got off at their own stops, no one put them right afterwards.

Weakness 2: generalisability. Every participant was riding one stretch of one city's subway in the middle of the day, so we can't be sure villagers, or people raised with different ideas about helping strangers, would act the same way. The victims were all men and only black or white, so we can't tell whether a woman, or someone of another ethnic group, would be helped in the same way.

How it would be marked: four points of equal depth, each with study details and an explanation, including the named issue: the top level (9–10 marks).

Common mistakes

- **Getting the victims wrong:** making the ill victim an old man, the drunk victim young, or some victims female; they were all men aged 26–35 (June 2023 report).
- **Giving the wrong group's features:** describing the victims or models when the **sample** was asked for, or the victims when the **models** or the **observers' measures** were asked for (November 2022, March 2023, June 2022 and June 2023 reports).
- **Saying the victim condition was random:** it alternated from day to day (June 2022 report).
- **Results with no comparison or no data:** "more help was given to the ill victim" isn't a meaningful comparison on its own, and data must be given when the question asks for it (November 2022 report).
- **Reversing the group-size result:** larger groups were **faster** to help, not more likely to help, and not slower (March 2021 report).
- **Describing diffusion of responsibility backwards** ("fewer people, less help") or as passive (March 2021 report).
- **Making up comments:** only comments recorded in the study count (June 2021 report).
- **Writing about the study when the background is asked,** or giving only Kitty Genovese (March 2026 report).
- **Calling it a natural experiment:** it was a field experiment, as the researchers manipulated the IVs (November 2022 report).
- **Defining a term with itself,** such as the right to withdraw or confidentiality, or confusing confidentiality with privacy (March 2023 report).
- **Strengths with no example from the study** when the question says "this study" (November 2022 report).
- **Missing the named issue** in the 10-mark evaluation, or writing a prepared essay that describes rather than evaluates (November 2021 and June 2022 reports).
- **Not linking an explicit difference** to the other research: the examiner can't make the link for you (June 2021 report).

How to get the marks

- **"State" and "identify":** a short exact answer (about 4450; alternated; white; aged 24–29). A list marked "definitive" only credits its own facts.
- **"Name two features":** two separate facts about the right group: sample, victims, models or observers.
- **"Describe"** a part of the procedure: one specific fact per mark, staying inside the start and end points given.
- **"Outline one result":** a meaningful comparison for 2 marks (more or less, faster or slower, than what), with data if the question asks for it.

- **"Outline one conclusion"**: a general claim about people, not a number, and not the one printed in the question.
- **"Explain"** a strength, weakness or reason: the point (1), then the Piliavin detail (1), then why it matters if there's a third mark.
- **"Suggest"** (a new scenario): link each point to something Piliavin found or did.
- **Two friends disagree**: choose one side, name the type of validity or generalisability, then give linked Piliavin evidence.
- **"Evaluate ... two strengths and two weaknesses"** (10 marks): four equal points with the named issue, at least two study details in each, no description.
- **Similarity and difference** (8 marks): state it, give an example from **each** study, and explain why it matters.

Quick check

1. A student writes: "The models were older than the victims." Use the two age ranges to explain whether this is right.
2. About 45% of the roughly 4450 passengers were black and 55% were white. Work out roughly how many passengers were in each group.
3. Say whether each is a result or a conclusion: (i) about 90% of the help came from men; (ii) in emergencies that need physical help, men tend to be the ones who act.
4. Why was it useful for the observers to count how many passengers were in the critical area on each trial?
5. 34 people left the critical area, spread over 21 trials. On those trials, how many people left on average? Give your answer to 1 decimal place.

Answers

1. Not quite. The models were aged 24–29 and the victims 26–35, so the ranges overlap from 26 to 29. The oldest victims (35) were $35 - 29 = 6$ years older than the oldest model, and the youngest model (24) was $26 - 24 = 2$ years younger than the youngest victim, so the models were, if anything, **younger**.
2. Black: $0.45 \times 4450 = 2002.5$, so about **2000**. White: $0.55 \times 4450 = 2447.5$, so about **2450**. (Check: $2002.5 + 2447.5 = 4450$.)
3. (i) A result (a measured finding about this study); (ii) a conclusion (a general claim about people).
4. It let the researchers test diffusion of responsibility: by comparing trials with more and fewer people nearby, they could see whether crowd size changed how often or how quickly the victim was helped (it turned out larger groups were quicker).
5. $34 \div 21 = 1.619 \dots$, so about **1.6** people per trial.

Past questions to try

- **9990/12/M/J/22 Q1**: the number of participants, how the victim condition was chosen, and two differences between the models.
- **9990/11/O/N/21 Q4**: what the victim did during a trial, then one weakness.
- **9990/12/O/N/22 Q8**: two sample features, then a two-friends question on generalisability.
- **9990/12/F/M/23 Q9**: the sample, then four ethical guidelines applied to the study.